

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2005-502322

(P2005-502322A)

(43) 公表日 平成17年1月27日(2005.1.27)

(51) Int.Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	4 B O 2 4
A 6 1 K 38/00	A 6 1 P 7/00	4 B O 6 3
A 6 1 K 38/22	C O 7 K 14/47	4 B O 6 4
A 6 1 P 7/00	C 1 2 N 1/15	4 B O 6 5
C O 7 K 14/47	C 1 2 N 1/19	4 C O 8 4
審査請求 未請求 予備審査請求 未請求 (全 317 頁) 最終頁に続く		

(21) 出願番号	特願2002-583449 (P2002-583449)	(71) 出願人	593052785
(86) (22) 出願日	平成14年4月19日 (2002.4.19)		ザ スクリップス リサーチ インスティテュート
(85) 翻訳文提出日	平成15年10月20日 (2003.10.20)		アメリカ合衆国 カリフォルニア州 92
(86) 国際出願番号	PCT/US2002/012465		037 ラ ジョラ ノース トーリー
(87) 国際公開番号	W02002/085923		パインズ ロード 10550
(87) 国際公開日	平成14年10月31日 (2002.10.31)	(74) 代理人	100064355
(31) 優先権主張番号	60/285,030		弁理士 川原田 一穂
(32) 優先日	平成13年4月19日 (2001.4.19)	(72) 発明者	ペーター・シュルツ
(33) 優先権主張国	米国 (US)		アメリカ合衆国 カリフォルニア州 92
(31) 優先権主張番号	60/355,514		037 ラ ジョラ ラ ジョラ ランチ
(32) 優先日	平成14年2月6日 (2002.2.6)		ロード 1650
(33) 優先権主張国	米国 (US)		
最終頁に続く			

(54) 【発明の名称】 非天然アミノ酸のインビボ組込み

(57) 【要約】

本発明は非天然アミノ酸を *in vivo* 組込むための方法と組成物を提供する。非天然アミノ酸を組込んだタンパク質を含む組成物も提供する。

【特許請求の範囲】

【請求項 1】

直交 tRNA (O-tRNA) と直交アミノアシル tRNA シンテターゼ (O-RS) を含む翻訳系を含む組成物であって、O-RS が翻訳系内で O-tRNA を少なくとも 1 種の非天然アミノ酸で優先的にアミノアシル化し、O-tRNA が少なくとも 1 種のセクターコドンを認識する前記組成物。

【請求項 2】

翻訳系が細胞を含む請求項 1 に記載の組成物。

【請求項 3】

翻訳系が細菌細胞を含む請求項 2 に記載の組成物。

10

【請求項 4】

翻訳系が大腸菌細胞を含む請求項 3 に記載の組成物。

【請求項 5】

翻訳系が始原菌細胞を含む請求項 2 に記載の組成物。

【請求項 6】

翻訳系が真核細胞を含む請求項 2 に記載の組成物。

【請求項 7】

翻訳系が酵母細胞、哺乳動物細胞、植物細胞又は昆虫細胞を含む請求項 6 に記載の組成物。

【請求項 8】

翻訳系が *in vitro* 翻訳系を含む請求項 1 に記載の組成物。

20

【請求項 9】

翻訳系が細胞抽出物を含む請求項 8 に記載の組成物。

【請求項 10】

O-tRNA が配列番号 1～3 及びその相補的ポリヌクレオチド配列から構成される群から選択されるポリヌクレオチド配列を含む核酸を含む請求項 1 に記載の組成物。

【請求項 11】

O-RS が、配列番号 4～34 から構成される群から選択されるアミノ酸配列を含むポリペプチドと、配列番号 35～66 及びその相補的ポリヌクレオチド配列から構成される群から選択されるポリヌクレオチド配列を含む核酸によりコードされるポリペプチドから構成される群から選択されるポリペプチドを含む請求項 1 に記載の組成物。

30

【請求項 12】

少なくとも 1 種の非天然アミノ酸が O-メチル-L-チロシン、L-3-(2-ナフチル)アラニン、3-メチルフェニルアラニン、O-4-アリル-L-チロシン、4-プロピル-L-チロシン、トリ-O-アセチル-GlcNAc β-セリン、L-Dopa、フッ素化フェニルアラニン、イソプロピル-L-フェニルアラニン、p-アジド-L-フェニルアラニン、p-アシル-L-フェニルアラニン、p-ベンゾイル-L-フェニルアラニン、L-ホスホセリン、ホスホノセリン、ホスホノチロシン、p-ヨードフェニルアラニン、p-ブロモフェニルアラニン、p-アミノ-L-フェニルアラニン、イソプロピル-L-フェニルアラニン、チロシナミノ酸の非天然類似体；グルタミナミノ酸の非天然類似体；フェニルアラニナミノ酸の非天然類似体；セリナミノ酸の非天然類似体；スレオニナミノ酸の非天然類似体；アルキル、アリール、アシル、アジド、シアノ、ハロ、ヒドラジン、ヒドラジド、ヒドロキシル、アルケニル、アルキニル、エーテル、チオール、スルホニル、セレノ、エステル、チオ酸、硼酸、ボロン酸、ホスホ、ホスホノ、ホスフィン、複素環、エノン、イミン、アルデヒド、ヒドロキシルアミン、ケトもしくはアミノ置換アミノ酸又はその任意組合せ；光架橋基をもつアミノ酸；スピンラベル付きアミノ酸；蛍光アミノ酸；新規官能基をもつアミノ酸；別の分子と共有又は非共有的に相互作用するアミノ酸；金属結合性アミノ酸；金属含有アミノ酸；放射性アミノ酸；フォトケージド及び／又は光異性化可能なアミノ酸；ビオチン又はビオチン類似体含有アミノ酸；グリコシル化又は糖鎖修飾アミノ酸；ケト含有アミノ酸；ポリエチレングリコール又はポリエ

40

50

ーテルを含むアミノ酸；重原子置換アミノ酸；化学分解性又は光分解性アミノ酸；延長側鎖をもつアミノ酸；毒性基を含むアミノ酸；糖鎖置換アミノ酸（例えば糖鎖置換セリン等）；炭素結合糖含有アミノ酸；レドックス活性アミノ酸； α -ヒドロキシ含有酸；アミノチオ酸含有アミノ酸； α , α ジ置換アミノ酸； β アミノ酸；及びプロリン以外の環状アミノ酸から構成される群から選択される請求項1に記載の組成物。

【請求項13】

少なくとも1種の非天然アミノ酸がO-メチル-L-チロシンである請求項12に記載の組成物。

【請求項14】

少なくとも1種の非天然アミノ酸がL-3-(2-ナフチル)アラニンである請求項12 10
に記載の組成物。

【請求項15】

少なくとも1種の非天然アミノ酸がアミノ-、イソプロピル-、又はO-アリル含有フェニルアラニン類似体である請求項12に記載の組成物。

【請求項16】

少なくとも1種のセクターコドンがナンセンスコドン、レアコドン又は4塩基コドンである請求項1に記載の組成物。

【請求項17】

少なくとも1種のセクターコドンがアンバーコドンである請求項16に記載の組成物。

【請求項18】

mtRNA_{CUA}^{Tyr}と突然変異体TyrRS(LWJ16)を含む大腸菌細胞であって、突然変異体TyrRS(LWJ16)が細胞内でmtRNA_{CUA}^{Tyr}をO-メチル-L-チロシンで優先的にアミノアシル化し、細胞がmtRNA_{CUA}^{Tyr}を使用してアンバーコドンを認識する前記大腸菌細胞。 20

【請求項19】

mtRNA_{CUA}^{Tyr}とSS12-TyrRSを含む大腸菌細胞であって、SS12-TyrRSが細胞内でmtRNA_{CUA}^{Tyr}をL-3-(2-ナフチル)アラニンで優先的にアミノアシル化し、細胞がmtRNA_{CUA}^{Tyr}を使用してアンバーコドンを認識する前記大腸菌細胞。

【請求項20】

少なくとも1種の非天然アミノ酸を含むタンパク質を含む組成物であって、タンパク質が少なくとも約100 μ g/Lの濃度で存在する前記組成物。 30

【請求項21】

タンパク質がサイトカイン、増殖因子、増殖因子受容体、インターフェロン、インターロイキン、炎症性分子、腫瘍遺伝子産物、ペプチドホルモン、シグナル伝達分子、ステロイドホルモン受容体、転写アクチベーター、転写サプレッサー、エリスロポエチン(EPO)、インスリン、ヒト成長ホルモン、上皮好中球活性化ペプチド-78、GRO α /MGSA、GRO β 、GRO γ 、MIP-1 α 、MIP-1 δ 、MCP-1、肝細胞増殖因子、インスリン様増殖因子、白血病阻害因子、オンコスタチンM、PD-ECSF、PDGF、プレイオトロピン、SCF、cキットリガンド、VEGEF、G-CSF、IL-1 40
、IL-2、IL-8、IGF-I、IGF-II、FGF(繊維芽細胞増殖因子)、PDGF、TNF、TGF- α 、TGF- β 、EGF(表皮増殖因子)、KGF(ケラチノサイト増殖因子)、SCF/cキット、CD40L/CD40、VLA-4/VCAM-1、ICAM-1/LFA-1、ヒアルリン/CD44、Mos、Ras、Raf、Met、p53、Tat、Fos、Myc、Jun、Myb、Rel、エストロゲン受容体、プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、LDL受容体及びコルチコステロン受容体から構成される群から選択される治療用タンパク質に相同である請求項20に記載の組成物。

【請求項22】

タンパク質が α 1アンチトリプシン、アンジオスタチン、抗血友病因子、抗体、アポリポ 50

タンパク質、アポタンパク質、心房性ナトリウム利尿因子、心房性ナトリウム利尿ポリペプチド、心房性ペプチド、C-X-Cケモカイン、T39765、NAP-2、ENA-78、Gro-a、Gro-b、Gro-c、IP-10、GCP-2、NAP-4、SDF-1、PF-4、MIG、カルシトニン、cキットリガンド、サイトカイン、CCケモカイン、単球化学誘引タンパク質-1、単球化学誘引タンパク質-2、単球化学誘引タンパク質-3、単球炎症性タンパク質-1 α 、単球炎症性タンパク質-1 β 、RANTES、I309、R83915、R91733、HCC1、T58847、D31065、T64262、CD40、CD40リガンド、Cキットリガンド、コラーゲン、コロニー刺激因子(CSF)、補体因子5a、補体阻害剤、補体受容体1、サイトカイン、上皮好中球活性化ペプチド-78、GRO α /MGSA、GRO β 、GRO γ 、MIP-1 α 、MIP-1 δ 、MCP-1、表皮増殖因子(EGF)、上皮好中球活性化ペプチド、エリスロポエチン(EPO)、表皮剥離毒素、IX因子、VII因子、VIII因子、X因子、繊維芽細胞増殖因子(FGF)、フィブリノーゲン、フィブロンectin、G-CSF、GM-CSF、グルコセレブロシダーゼ、ゴナドトロピン、増殖因子、増殖因子受容体、ヘッジホッグタンパク質、ヘモグロビン、肝細胞増殖因子(HGF)、ヒルジン、ヒト血清アルブミン、ICAM-1、ICAM-1受容体、LFA-1、LFA-1受容体、インスリン、インスリン様増殖因子(IGF)、IGF-I、IGF-II、インターフェロン、IFN- α 、IFN- β 、IFN- γ 、インターロイキン、IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-11、IL-12、ケラチノサイト増殖因子(KGF)、ラクトフェリン、白血病阻害因子、ルシフェラーゼ、ニュールチュリン、好中球阻害因子(NIF)、オンコスタチンM、骨形成タンパク質、腫瘍遺伝子産物、副甲状腺ホルモン、PD-ECSF、PDGF、ペプチドホルモン、ヒト増殖ホルモン、プレイオトロピン、プロテインA、プロテインG、発熱外毒素A、B又はC、リラキシン、レニン、SCF、可溶性補体受容体I、可溶性I-CAM1、可溶性インターロイキン受容体、可溶性TNF受容体、ソマトメジン、ソマトスタチン、ソマトトロピン、ストレプトキナーゼ、スーパー抗原、ブドウ球菌エンテロトキシン、SEA、SEB、SEC1、SEC2、SEC3、SED、SEE、ステロイドホルモン受容体、スーパーオキシドジスムターゼ、毒素性ショック症候群毒素、チモシン α 1、組織プラスミノーゲンアクチベーター、腫瘍増殖因子(TGF)、TGF- α 、TGF- β 、腫瘍壊死因子、腫瘍壊死因子 α 、腫瘍壊死因子 β 、腫瘍壊死因子受容体(TNFR)、VLA-4タンパク質、VCAM-1タンパク質、血管内皮増殖因子(VEGEF)、ウロキナーゼ、Mos、Ras、Raf、Met、p53、Tat、Fos、Myc、Jun、Myb、Rel、エストロゲン受容体、プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、LDL受容体及びコルチコステロン受容体から構成される群から選択される治療用タンパク質に相同である請求項20に記載の組成物。

【請求項23】

タンパク質が治療用タンパク質に少なくとも約50%又は少なくとも約75%相同である請求項22に記載の組成物。

【請求項24】

治療用タンパク質がエリスロポエチン(EPO)である請求項22に記載の組成物。

【請求項25】

タンパク質が少なくとも2種の非天然アミノ酸を含む請求項20に記載の組成物。

【請求項26】

タンパク質が少なくとも3種の非天然アミノ酸を含む請求項20に記載の組成物。

【請求項27】

タンパク質が少なくとも4種の非天然アミノ酸を含む請求項20に記載の組成物。

【請求項28】

タンパク質が少なくとも5種又は6種以上の非天然アミノ酸を含む請求項20に記載の組成物。

10

20

30

40

50

【請求項 29】

組成物が少なくとも約 100 μg のタンパク質を含む請求項 20 に記載の組成物。

【請求項 30】

タンパク質が DHFR であり、少なくとも 1 種の非天然アミノ酸が O-メチル-L-チロシンと L-3-(2-ナフチル)アラニンから構成される群から選択される請求項 20 に記載の組成物。

【請求項 31】

少なくとも 1 種の非天然アミノ酸を含むタンパク質を含む組成物であって、タンパク質がサイトカイン、増殖因子、増殖因子受容体、インターフェロン、インターロイキン、炎症性分子、腫瘍遺伝子産物、ペプチドホルモン、シグナル伝達分子、ステロイドホルモン受容体、転写アクチベーター、転写サプレッサー、エリスロポエチン (EPO)、インスリン、ヒト成長ホルモン、上皮好中球活性化ペプチド-78、GRO α /MGSA、GRO β 、GRO γ 、MIP-1 α 、MIP-1 δ 、MCP-1、肝細胞増殖因子、インスリン様増殖因子、白血病阻害因子、オンコスタチンM、PD-ECSF、PDGF、プレイオトロピン、SCF、cキットリガンド、VEGEF、G-CSF、IL-1、IL-2、IL-8、IGF-I、IGF-II、FGF (繊維芽細胞増殖因子)、PDGF、TNF、TGF- α 、TGF- β 、EGF (表皮増殖因子)、KGF (ケラチノサイト増殖因子)、SCF/cキット、CD40L/CD40、VLA-4/VCAM-1、ICAM-1/LFA-1、ヒアルリン/CD44、Mos、Ras、Raf、Met、p53、Tat、Fos、Myc、Jun、Myb、Rel、エストロゲン受容体、プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、LDL受容体及びコレステロン受容体から構成される群から選択される治療用タンパク質に相同である前記組成物。

【請求項 32】

少なくとも 1 種の非天然アミノ酸を含むタンパク質を含む組成物であって、タンパク質が α 1アンチトリプシン、アンジオスタチン、抗血友病因子、抗体、アポリポタンパク質、アポタンパク質、心房性ナトリウム利尿因子、心房性ナトリウム利尿ポリペプチド、心房性ペプチド、C-X-Cケモカイン、T39765、NAP-2、ENA-78、Gro-a、Gro-b、Gro-c、IP-10、GCP-2、NAP-4、SDF-1、PF-4、MIG、カルシトニン、cキットリガンド、サイトカイン、CCケモカイン、単球化学誘引タンパク質-1、単球化学誘引タンパク質-2、単球化学誘引タンパク質-3、単球炎症性タンパク質-1 α 、単球炎症性タンパク質-1 β 、RANTES、I309、R83915、R91733、HCC1、T58847、D31065、T64262、CD40、CD40リガンド、Cキットリガンド、コラーゲン、コロニー刺激因子 (CSF)、補体因子5a、補体阻害剤、補体受容体1、サイトカイン、上皮好中球活性化ペプチド-78、GRO α /MGSA、GRO β 、GRO γ 、MIP-1 α 、MIP-1 δ 、MCP-1、表皮増殖因子 (EGF)、上皮好中球活性化ペプチド、エリスロポエチン (EPO)、表皮剥離毒素、IX因子、VII因子、VII因子、X因子、繊維芽細胞増殖因子 (FGF)、フィブリノーゲン、フィブロンectin、G-CSF、GM-CSF、グルコセレブロシダーゼ、ゴナドトロピン、増殖因子、増殖因子受容体、ヘッジホッグタンパク質、ヘモグロビン、肝細胞増殖因子 (HGF)、ヒルジン、ヒト血清アルブミン、ICAM-1、ICAM-1受容体、LFA-1、LFA-1受容体、インスリン、インスリン様増殖因子 (IGF)、IGF-I、IGF-II、インターフェロン、IFN- α 、IFN- β 、IFN- γ 、インターロイキン、IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-11、IL-12、ケラチノサイト増殖因子 (KGF)、ラクトフェリン、白血病阻害因子、ルシフェラーゼ、ニュールチュリン、好中球阻害因子 (NIF)、オンコスタチンM、骨形成タンパク質、腫瘍遺伝子産物、副甲状腺ホルモン、PD-ECSF、PDGF、ペプチドホルモン、ヒト増殖ホルモン、プレイオトロピン、プロテインA、プロテインG、発熱外毒素A、B又はC、リラキシン、レニン、SCF、可溶性補体受容体I、可溶性I-

C A M 1、可溶性インターロイキン受容体、可溶性 T N F 受容体、ソマトメジン、ソマトスタチン、ソマトトロピン、ストレプトキナーゼ、スーパー抗原、ブドウ球菌エンテロトキシン、S E A、S E B、S E C 1、S E C 2、S E C 3、S E D、S E E、ステロイドホルモン受容体、スーパーオキシドジスムターゼ、毒素性ショック症候群毒素、チモシン α 1、組織プラスミノゲンアクチベーター、腫瘍増殖因子 (T G F)、T G F - α 、T G F - β 、腫瘍壊死因子、腫瘍壊死因子 α 、腫瘍壊死因子 β 、腫瘍壊死因子受容体 (T N F R)、V L A - 4 タンパク質、V C A M - 1 タンパク質、血管内皮増殖因子 (V E G E F)、ウロキナーゼ、M o s、R a s、R a f、M e t、p 5 3、T a t、F o s、M y c、J u n、M y b、R e l、エストロゲン受容体、プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、L D L 受容体及びコルチコステロン受容体から構成される群から選択される治療用タンパク質に相同である前記組成物。 10

【請求項 3 3】

タンパク質が少なくとも 2 種の非天然アミノ酸を含む請求項 3 2 に記載の組成物。

【請求項 3 4】

タンパク質が少なくとも 3 種の非天然アミノ酸を含む請求項 3 2 に記載の組成物。

【請求項 3 5】

タンパク質が少なくとも 4 種の非天然アミノ酸を含む請求項 3 2 に記載の組成物。

【請求項 3 6】

タンパク質が少なくとも 5 種又は 6 種以上の非天然アミノ酸を含む請求項 3 2 に記載の組成物。 20

【請求項 3 7】

タンパク質が治療用タンパク質に少なくとも約 5 0 % 又は少なくとも約 7 5 % 相同である請求項 3 2 に記載の組成物。

【請求項 3 8】

治療用タンパク質がエリスロポエチン (E P O) である請求項 3 2 に記載の組成物。

【請求項 3 9】

タンパク質と医薬的に許容可能な賦形剤を含む組成物であって、タンパク質が少なくとも 1 種の非天然アミノ酸を含む前記組成物。

【請求項 4 0】

タンパク質が少なくとも 2 種の非天然アミノ酸を含む請求項 3 9 に記載の組成物。 30

【請求項 4 1】

タンパク質が少なくとも 3 種の非天然アミノ酸を含む請求項 3 9 に記載の組成物。

【請求項 4 2】

タンパク質が少なくとも 4 種の非天然アミノ酸を含む請求項 3 9 に記載の組成物。

【請求項 4 3】

タンパク質が少なくとも 5 種又は 6 種以上の非天然アミノ酸を含む請求項 3 9 に記載の組成物。

【請求項 4 4】

タンパク質がサイトカイン、増殖因子、増殖因子受容体、インターフェロン、インターロイキン、炎症性分子、腫瘍遺伝子産物、ペプチドホルモン、シグナル伝達分子、ステロイドホルモン受容体、転写アクチベーター、転写サプレッサー、エリスロポエチン (E P O)、インスリン、ヒト成長ホルモン、上皮好中球活性化ペプチド - 7 8、G R O α / M G S A、G R O β 、G R O γ 、M I P - 1 α 、M I P - 1 δ 、M C P - 1、肝細胞増殖因子、インスリン様増殖因子、白血病阻害因子、オンコスタチン M、P D - E C S F、P D G F、プレイオトロピン、S C F、c キットリガンド、V E G E F、G - C S F、I L - 1、I L - 2、I L - 8、I G F - I、I G F - I I、F G F (繊維芽細胞増殖因子)、P D G F、T N F、T G F - α 、T G F - β 、E G F (表皮増殖因子)、K G F (ケラチノサイト増殖因子)、S C F / c キット、C D 4 0 L / C D 4 0、V L A - 4 / V C A M - 1、I C A M - 1 / L F A - 1、ヒアルリン / C D 4 4、M o s、R a s、R a f、M e t、p 5 3、T a t、F o s、M y c、J u n、M y b、R e l、エストロゲン受容体、 40 50

プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、LDL受容体及びコルチコステロン受容体から構成される群から選択される治療用タンパク質に相同である請求項39に記載の組成物。

【請求項45】

タンパク質が $\alpha 1$ アンチトリプシン、アンジオスタチン、抗血友病因子、抗体、アポリポタンパク質、アポタンパク質、心房性ナトリウム利尿因子、心房性ナトリウム利尿ポリペプチド、心房性ペプチド、C-X-Cケモカイン、T39765、NAP-2、ENA-78、Gro-a、Gro-b、Gro-c、IP-10、GCP-2、NAP-4、SDF-1、PF-4、MIG、カルシトニン、cキットリガンド、サイトカイン、CCケモカイン、単球化学誘引タンパク質-1、単球化学誘引タンパク質-2、単球化学誘引タンパク質-3、単球炎症性タンパク質-1 α 、単球炎症性タンパク質-1 β 、RANTES、I309、R83915、R91733、HCC1、T58847、D31065、T64262、CD40、CD40リガンド、Cキットリガンド、コラーゲン、コロニー刺激因子(CSF)、補体因子5a、補体阻害剤、補体受容体1、サイトカイン、上皮好中球活性化ペプチド-78、GRO α /MGSA、GRO β 、GRO γ 、MIP-1 α 、MIP-1 δ 、MCP-1、表皮増殖因子(EGF)、上皮好中球活性化ペプチド、エリスロポエチン(EPO)、表皮剥離毒素、IX因子、VII因子、VIII因子、X因子、繊維芽細胞増殖因子(FGF)、フィブリノーゲン、フィブロネクチン、G-CSF、GM-CSF、グルコセレブロシダーゼ、ゴナドトロピン、増殖因子、増殖因子受容体、ヘッジホッグタンパク質、ヘモグロビン、肝細胞増殖因子(HGF)、ヒルジン、ヒト血清アルブミン、ICAM-1、ICAM-1受容体、LFA-1、LFA-1受容体、インスリン、インスリン様増殖因子(IGF)、IGF-I、IGF-II、インターフェロン、IFN- α 、IFN- β 、IFN- γ 、インターロイキン、IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-11、IL-12、ケラチノサイト増殖因子(KGF)、ラクトフェリン、白血病阻害因子、ルシフェラーゼ、ニールチュリン、好中球阻害因子(NIF)、オンコスタチンM、骨形成タンパク質、腫瘍遺伝子産物、副甲状腺ホルモン、PD-ECSF、PDGF、ペプチドホルモン、ヒト増殖ホルモン、プレイオトロピン、プロテインA、プロテインG、発熱外毒素A、B又はC、リラキシン、レニン、SCF、可溶性補体受容体I、可溶性ICAM1、可溶性インターロイキン受容体、可溶性TNF受容体、ソマトメジン、ソマトスタチン、ソマトトロピン、ストレプトキナーゼ、スーパー抗原、ブドウ球菌エンテロトキシン、SEA、SEB、SEC1、SEC2、SEC3、SED、SEE、ステロイドホルモン受容体、スーパーオキシドジスムターゼ、毒素性ショック症候群毒素、チモシン $\alpha 1$ 、組織プラスミノゲンアクチベーター、腫瘍増殖因子(TGF)、TGF- α 、TGF- β 、腫瘍壊死因子、腫瘍壊死因子 α 、腫瘍壊死因子 β 、腫瘍壊死因子受容体(TNFR)、VLA-4タンパク質、VCAM-1タンパク質、血管内皮増殖因子(VEGEF)、ウロキナーゼ、Mos、Ras、Raf、Met、p53、Tat、Fos、Myc、Jun、Myb、Rel、エストロゲン受容体、プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、LDL受容体及びコルチコステロン受容体から構成される群から選択される治療用タンパク質に相同である請求項39に記載の組成物。

【請求項46】

タンパク質が治療用タンパク質に少なくとも約50%又は少なくとも約75%相同である請求項45に記載の組成物。

【請求項47】

治療用タンパク質がエリスロポエチン(EPO)である請求項45に記載の組成物。

【請求項48】

少なくとも1種の非天然アミノ酸を含む少なくとも1種のタンパク質を翻訳系で生産する方法であって、

少なくとも1種のセクターコドンを含み、前記少なくとも1種のタンパク質をコードす

10

20

30

40

50

る少なくとも1種の核酸を翻訳系に加える段階と、
直交tRNA（O-tRNA）（但し、O-tRNAは翻訳系で機能し、O-tRNAは少なくとも1種のセクターコドンを認識する）を翻訳系に加える段階と、
直交アミノアシルtRNAシンテターゼ（O-RS）（但し、O-RSは翻訳系内でO-tRNAを少なくとも1種の非天然アミノ酸で優先的にアミノアシル化する）を翻訳系に加える段階と、
少なくとも1種の非天然アミノ酸を翻訳系に加え、少なくとも1種の非天然アミノ酸を含む少なくとも1種のタンパク質を翻訳系で生産する段階を含む前記方法。

【請求項49】

翻訳系が細胞を含む請求項48に記載の方法。

10

【請求項50】

翻訳系が細菌細胞を含む請求項49に記載の方法。

【請求項51】

翻訳系が大腸菌細胞を含む請求項50に記載の方法。

【請求項52】

翻訳系が始原菌細胞を含む請求項49に記載の方法。

【請求項53】

翻訳系が真核細胞を含む請求項49に記載の方法。

【請求項54】

翻訳系が酵母細胞、哺乳動物細胞、植物細胞又は昆虫細胞を含む請求項53に記載の方法。

20

【請求項55】

請求項49に記載の方法により生産した少なくとも1種の非天然アミノ酸を含むタンパク質であって、タンパク質が細胞依存的にプロセッシング及び改変されている前記タンパク質。

【請求項56】

タンパク質がサイトカイン、増殖因子、増殖因子受容体、インターフェロン、インターロイキン、炎症性分子、腫瘍遺伝子産物、ペプチドホルモン、シグナル伝達分子、ステロイドホルモン受容体、転写アクチベーター、転写サプレッサー、エリスロポエチン（EPO）、インスリン、ヒト成長ホルモン、上皮好中球活性化ペプチド-78、GRO α /MGS α 、GRO β 、GRO γ 、MIP-1 α 、MIP-1 δ 、MCP-1、肝細胞増殖因子、インスリン様増殖因子、白血病阻害因子、オンコスタチンM、PD-ECSF、PDGF、プレイオトロピン、SCF、cキットリガンド、VEGEF、G-CSF、IL-1、IL-2、IL-8、IGF-I、IGF-II、FGF（繊維芽細胞増殖因子）、PDGF、TNF、TGF- α 、TGF- β 、EGF（表皮増殖因子）、KGF（ケラチノサイト増殖因子）、SCF/cキット、CD40L/CD40、VLA-4/VCAM-1、ICAM-1/LFA-1、ヒアルリン/CD44、Mos、Ras、Raf、Met、p53、Tat、Fos、Myc、Jun、Myb、Rel、エストロゲン受容体、プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、LDL受容体及びコルチコステロン受容体から構成される群から選択される治療用タンパク質に相同である請求項55に記載のタンパク質。

30

40

【請求項57】

タンパク質が α 1アンチトリプシン、アンジオスタチン、抗血友病因子、抗体、アポリポタンパク質、アポタンパク質、心房性ナトリウム利尿因子、心房性ナトリウム利尿ポリペプチド、心房性ペプチド、C-X-Cケモカイン、T39765、NAP-2、ENA-78、Gro-a、Gro-b、Gro-c、IP-10、GCP-2、NAP-4、SDF-1、PF-4、MIG、カルシトニン、cキットリガンド、サイトカイン、CCケモカイン、単球化学誘引タンパク質-1、単球化学誘引タンパク質-2、単球化学誘引タンパク質-3、単球炎症性タンパク質-1 α 、単球炎症性タンパク質-1 β 、RANTE、I309、R83915、R91733、HCC1、T58847、D31065、

50

T 6 4 2 6 2、C D 4 0、C D 4 0 リガンド、C キットリガンド、コラーゲン、コロニー刺激因子 (C S F)、補体因子 5 a、補体阻害剤、補体受容体 1、サイトカイン、上皮好中球活性化ペプチド-7 8、G R O α / M G S A、G R O β、G R O γ、M I P-1 α、M I P-1 δ、M C P-1、表皮増殖因子 (E G F)、上皮好中球活性化ペプチド、エリスロポエチン (E P O)、表皮剥離毒素、I X 因子、V I I 因子、V I I I 因子、X 因子、繊維芽細胞増殖因子 (F G F)、フィブリノーゲン、フィブロネクチン、G-C S F、G M-C S F、グルコセレブロシダーゼ、ゴナドトロピン、増殖因子、増殖因子受容体、ヘッジホッグタンパク質、ヘモグロビン、肝細胞増殖因子 (H G F)、ヒルジン、ヒト血清アルブミン、I C A M-1、I C A M-1 受容体、L F A-1、L F A-1 受容体、インスリン、インスリン様増殖因子 (I G F)、I G F-I、I G F-I I、インターフェロン、I F N-α、I F N-β、I F N-γ、インターロイキン、I L-1、I L-2、I L-3、I L-4、I L-5、I L-6、I L-7、I L-8、I L-9、I L-10、I L-11、I L-12、ケラチノサイト増殖因子 (K G F)、ラクトフェリン、白血病阻害因子、ルシフェラーゼ、ニュールチュリン、好中球阻害因子 (N I F)、オンコスタチン M、骨形成タンパク質、腫瘍遺伝子産物、副甲状腺ホルモン、P D-E C S F、P D G F、ペプチドホルモン、ヒト増殖ホルモン、プレイオトロピン、プロテイン A、プロテイン G、発熱外毒素 A、B 又は C、リラキシン、レニン、S C F、可溶性補体受容体 I、可溶性 I-C A M 1、可溶性インターロイキン受容体、可溶性 T N F 受容体、ソマトメジン、ソマトスタチン、ソマトトロピン、ストレプトキナーゼ、スーパー抗原、ブドウ球菌エンテロトキシン、S E A、S E B、S E C 1、S E C 2、S E C 3、S E D、S E E、ステロイドホルモン受容体、スーパーオキシドジスムターゼ、毒素性ショック症候群毒素、チモシン α 1、組織プラスミノゲンアクチベーター、腫瘍増殖因子 (T G F)、T G F-α、T G F-β、腫瘍壊死因子、腫瘍壊死因子 α、腫瘍壊死因子 β、腫瘍壊死因子受容体 (T N F R)、V L A-4 タンパク質、V C A M-1 タンパク質、血管内皮増殖因子 (V E G E F)、ウロキナーゼ、M o s、R a s、R a f、M e t、p 5 3、T a t、F o s、M y c、J u n、M y b、R e l、エストロゲン受容体、プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、L D L 受容体及びコルチコステロン受容体から構成される群から選択される治療用タンパク質に相同である請求項 5 5 に記載のタンパク質。

【請求項 5 8】 30

タンパク質が治療用タンパク質に少なくとも約 5 0 % 又は少なくとも約 7 5 % 相同である請求項 5 7 に記載のタンパク質。

【請求項 5 9】

治療用タンパク質がエリスロポエチン (E P O) である請求項 5 7 に記載のタンパク質。

【請求項 6 0】

タンパク質が少なくとも 2 種の非天然アミノ酸を含む請求項 5 5 に記載のタンパク質。

【請求項 6 1】

タンパク質が少なくとも 3 種の非天然アミノ酸を含む請求項 5 5 に記載のタンパク質。

【請求項 6 2】

タンパク質が少なくとも 4 種の非天然アミノ酸を含む請求項 5 5 に記載のタンパク質。 40

【請求項 6 3】

タンパク質が少なくとも 5 種又は 6 種以上の非天然アミノ酸を含む請求項 5 5 に記載のタンパク質。

【請求項 6 4】

翻訳系が *i n v i v o* 翻訳系を含む請求項 4 8 に記載の方法。

【請求項 6 5】

翻訳系が細胞抽出物を含む請求項 6 4 に記載の方法。

【請求項 6 6】

少なくとも 1 種のタンパク質がクロラムフェニコールアセチルトランスフェラーゼ (C A T) の A s p 1 1 2 T A G 突然変異体、マウスジヒドロ葉酸レダクターゼ (D H F R) の 50

T y r 1 6 3 T A G 突然変異体、又は C O O H H i s 6 タグを付けたマウスジヒドロ葉酸レダクターゼ (D H F R) の T y r 1 6 3 T A G 突然変異体である請求項 4 8 に記載の方法。

【請求項 6 7】

非天然アミノ酸が O-メチル-L-チロシン、L-3-(2-ナフチル)アラニン、3-メチルフェニルアラニン、O-4-アリル-L-チロシン、4-プロピル-L-チロシン、トリ-O-アセチル-G l c N A c β-セリン、L-D o p a、フッ素化フェニルアラニン、イソプロピル-L-フェニルアラニン、p-アジド-L-フェニルアラニン、p-アシル-L-フェニルアラニン、p-ベンゾイル-L-フェニルアラニン、L-ホスホセリン、ホスホノセリン、ホスホノチロシン、p-ヨードフェニルアラニン、p-ブロモフェニルアラニン、p-アミノ-L-フェニルアラニン、イソプロピル-L-フェニルアラニン、チロシンアミノ酸の非天然類似体；グルタミンアミノ酸の非天然類似体；フェニルアラニンアミノ酸の非天然類似体；セリンアミノ酸の非天然類似体；スレオニンアミノ酸の非天然類似体；アルキル、アリール、アシル、アジド、シアノ、ハロ、ヒドラジン、ヒドラジド、ヒドロキシル、アルケニル、アルキニル、エーテル、チオール、スルホニル、セレノ、エステル、チオ酸、硼酸、ボロン酸、ホスホ、ホスホノ、ホスフィン、複素環、エノン、イミン、アルデヒド、ヒドロキシルアミン、ケトもしくはアミノ置換アミノ酸又はその任意組合せ；光架橋基をもつアミノ酸；スピンラベル付きアミノ酸；蛍光アミノ酸；新規官能基をもつアミノ酸；別の分子と共有又は非共有的に相互作用するアミノ酸；金属結合性アミノ酸；金属含有アミノ酸；放射性アミノ酸；フォトケージド及び／又は光異性化可能なアミノ酸；ビオチン又はビオチン類似体含有アミノ酸；グリコシル化又は糖鎖修飾アミノ酸；ケト含有アミノ酸；ポリエチレングリコール又はポリエーテルを含むアミノ酸；重原子置換アミノ酸；化学分解性又は光分解性アミノ酸；延長側鎖をもつアミノ酸；毒性基を含むアミノ酸；糖鎖置換アミノ酸（例えば糖鎖置換セリン等）；炭素結合糖含有アミノ酸；レドックス活性アミノ酸；α-ヒドロキシ含有酸；アミノチオ酸含有アミノ酸；α, αジ置換アミノ酸；βアミノ酸；及びプロリン以外の環状アミノ酸から構成される群から選択される請求項 4 8 に記載の方法。

10

20

【請求項 6 8】

少なくとも 1 種の非天然アミノ酸が O-メチル-L-チロシンである請求項 4 8 に記載の方法。

30

【請求項 6 9】

少なくとも 1 種の非天然アミノ酸が L-3-(2-ナフチル)アラニンである請求項 4 8 に記載の方法。

【請求項 7 0】

少なくとも 1 種の非天然アミノ酸がアミノ、イソプロピル、又は O-アリル含有フェニルアラニン類似体である請求項 4 8 に記載の方法。

【請求項 7 1】

O-t R N A が配列番号 1 ~ 3 及びその相補的ポリヌクレオチド配列から構成される群から選択されるポリヌクレオチド配列を含む核酸を含む請求項 4 8 に記載の方法。

【請求項 7 2】

少なくとも 1 種のセクターコドンがナンセンスコドン、レアコドン又は 4 塩基コドンである請求項 4 8 に記載の方法。

40

【請求項 7 3】

少なくとも 1 種のセクターコドンがアンバーコドンである請求項 7 2 に記載の方法。

【請求項 7 4】

O-R S が O-t R N A を O-メチル-L-チロシンで優先的にアミノアシル化する請求項 4 8 に記載の方法。

【請求項 7 5】

O-R S が O-t R N A をアミノ、イソプロピル、又は O-アリル含有フェニルアラニン類似体で優先的にアミノアシル化する、請求項 4 8 に記載の方法。

50

【請求項 76】

Ｏ－ＲＳが、配列番号４～３４から構成される群から選択されるアミノ酸配列を含むポリペプチドと、配列番号３５～６６及びその相補的ポリヌクレオチド配列から構成される群から選択されるポリヌクレオチド配列を含む核酸によりコードされるポリペプチドから構成される群から選択されるポリペプチドを含む請求項４８に記載の方法。

【請求項 77】

非天然アミノ酸を外部から加える請求項４８に記載の方法。

【請求項 78】

翻訳系が細胞であり、非天然アミノ酸が細胞により生合成される請求項４８に記載の方法。

10

【請求項 79】

少なくとも１種のＯ－メチル－Ｌ－チロシンを含む少なくとも１種のタンパク質を大腸菌細胞で生産する方法であって、アンバーコドンを含み、前記少なくとも１種のタンパク質をコードする少なくとも１種の核酸を翻訳系に加える段階と、

$mtRNA_{CUA}^{Tyr}$ （但し、 $mtRNA_{CUA}^{Tyr}$ は細胞で機能し、 $mtRNA_{CUA}^{Tyr}$ はアンバーコドンを認識する）を翻訳系に加える段階と、

突然変異体 $TyrRS(LWJ16)$ （但し、突然変異体 $TyrRS(LWJ16)$ は細胞内で $mtRNA_{CUA}^{Tyr}$ をＯ－メチル－Ｌ－チロシンでアミノアシル化する）を翻訳系に加える段階と、

20

Ｏ－メチル－Ｌ－チロシンを細胞に加え、Ｏ－メチル－Ｌ－チロシンを含む少なくとも１種のタンパク質を細胞内で生産する段階を含む前記方法。

【請求項 80】

少なくとも１種のＬ－３－（２－ナフチル）アラニンを含む少なくとも１種のタンパク質を大腸菌細胞で生産する方法であって、アンバーコドンを含み、前記少なくとも１種のタンパク質をコードする少なくとも１種の核酸を翻訳系に加える段階と、

$mtRNA_{CUA}^{Tyr}$ （但し、 $mtRNA_{CUA}^{Tyr}$ は細胞で機能し、 $mtRNA_{CUA}^{Tyr}$ はアンバーコドンを認識する）を細胞に加える段階と、

$SS12-TyrRS$ （但し、 $SS12-TyrRS$ は細胞内で $mtRNA_{CUA}^{Tyr}$ をＬ－３－（２－ナフチル）アラニンでアミノアシル化する）を細胞に加える段階と、

30

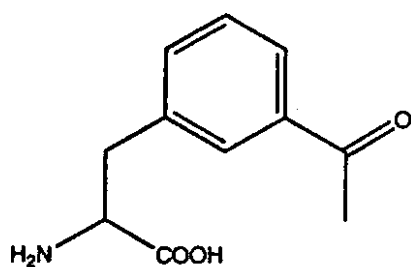
Ｌ－３－（２－ナフチル）アラニンを細胞に加え、Ｌ－３－（２－ナフチル）アラニンを含む少なくとも１種のタンパク質を細胞内で生産する段階を含む前記方法。

【請求項 81】

式 I :

【化 1】

I



40

に示す非天然アミノ酸を含む組成物。

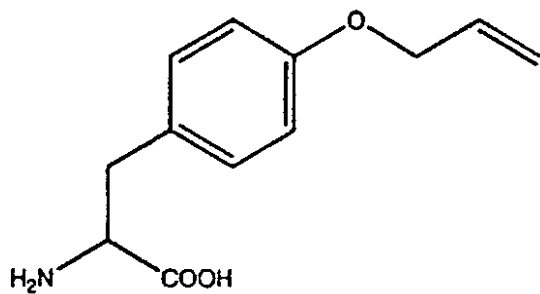
【請求項 82】

50

式 I I :

【化 2】

II



10

に示す非天然アミノ酸を含む組成物。

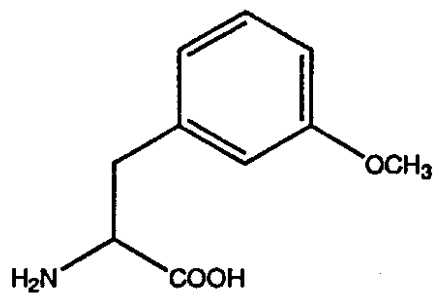
【請求項 8 3】

式 I I I :

20

【化 3】

III



30

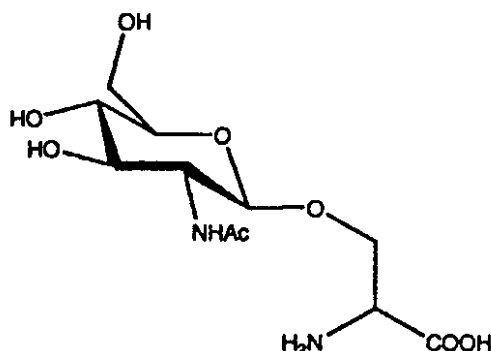
に示す非天然アミノ酸を含む組成物。

【請求項 8 4】

式 I V :

【化 4】

IV



10

に示す非天然アミノ酸を含む組成物。

【請求項 8 5】

更に直交 tRNA を含む請求項 8 1、請求項 8 2、請求項 8 3 又は請求項 8 4 に記載の組成物。 20

【請求項 8 6】

非天然アミノ酸が直交 tRNA に共有結合している請求項 8 5 に記載の組成物。

【請求項 8 7】

非天然アミノ酸がアミノアシル結合を介して直交 tRNA に共有結合している請求項 8 5 に記載の組成物。

【請求項 8 8】

非天然アミノ酸が直交 tRNA の末端リボース糖の 3' OH 又は 2' OH に共有結合している請求項 8 5 に記載の組成物。

【請求項 8 9】

更に直交アミノアシル tRNA シンテターゼを含む請求項 8 1、請求項 8 2、請求項 8 3 又は請求項 8 4 に記載の組成物。 30

【請求項 9 0】

非天然アミノ酸が直交アミノアシル tRNA シンテターゼに水素結合している請求項 8 9 に記載の組成物。

【請求項 9 1】

請求項 8 1、請求項 8 2、請求項 8 3 又は請求項 8 4 に記載の非天然アミノ酸を含むタンパク質。

【請求項 9 2】

請求項 8 1、請求項 8 2、請求項 8 3 又は請求項 8 4 に記載の非天然アミノ酸を含む細胞。 40

【請求項 9 3】

4-アリル-L-チロシンの合成方法であって、

(a) 保護チロシンを臭化アリルと反応させて保護アリルチロシンを得る段階と、
(b) 保護アリルチロシンを脱保護して 4-アリル-L-チロシンを生産する段階を含む前記方法。

【請求項 9 4】

水素化ナトリウムと DMF の存在下に臭化アリルと保護チロシンを反応させる請求項 9 3 に記載の方法。

【請求項 9 5】

50

保護チロシンが N B o c 保護チロシンを含む請求項 9 3 に記載の方法。

【請求項 9 6】

段階 (b) において保護アリルチロシンを酸性溶液中で脱保護する請求項 9 5 に記載の方法。

【請求項 9 7】

段階 (b) において保護アリルチロシンを塩酸とジオキサンの存在下に脱保護する請求項 9 5 に記載の方法。

【請求項 9 8】

4-アリル-L-チロシンをエタノール又はジクロロメタンで抽出する段階を更に含む請求項 9 3 に記載の方法。

10

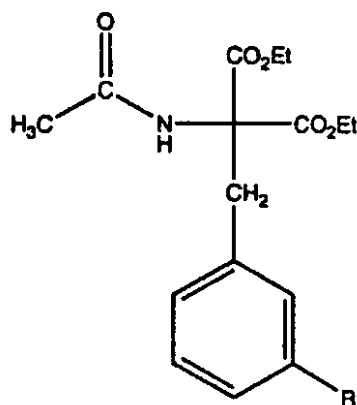
【請求項 9 9】

メタ置換フェニルアラニン類似体の合成方法であって、

(a) アセトアミドマロン酸ジエチルとメタ置換臭化ベンジルを縮合させ、式 V :

【化 5】

V



20

30

の化合物を得る段階と、

(b) 式 V の化合物を加水分解してメタ置換フェニルアラニン類似体を得る段階を含む前記方法。

【請求項 1 0 0】

R がケト、アセチル又はメトキシ基を含む請求項 9 9 に記載の方法。

【請求項 1 0 1】

メタ置換フェニルアラニン類似体が 3-メトキシフェニルアラニン又は 3-アセチルフェニルアラニンを含む請求項 9 9 に記載の方法。

【請求項 1 0 2】

式 V の化合物が 1-(3-ブロモメチルフェニル)エタノンを含む請求項 9 9 に記載の方法。

40

【請求項 1 0 3】

N-ブロモスクシンイミドを 3-メチルアセトフェノンと反応させて臭素化物を生成し、臭素化物をヘキサン溶液中で結晶化させることにより、メタ置換臭化ベンジルを得る段階を更に含む請求項 9 9 に記載の方法。

【請求項 1 0 4】

グリコシル化アミノ酸の *in vivo* 製造方法であって、N-アセチルガラクトサミニダーゼ、トランスグリコシラーゼ又はセリングリコシルヒドロラーゼの遺伝子を含むプラスミドで細胞を形質転換することを含む前記方法。

50

【請求項 105】

p-アミノフェニルアラニンの製造方法であって、

- (a) コリスミン酸を4-アミノ-4-デオキシコリスミン酸に酵素変換する段階と、
- (b) 4-アミノ-4-デオキシコリスミン酸を4-アミノ-4-デオキシプレフェン酸に酵素変換する段階と、
- (c) 4-アミノ-4-デオキシプレフェン酸をp-アミノフェニルピルビン酸に酵素変換する段階と、
- (d) p-アミノフェニルピルビン酸をp-アミノフェニルアラニンに酵素変換する段階を含む前記方法。

【請求項 106】

段階 (a) においてコリスミン酸を4-アミノ-4-デオキシコリスミン酸シンターゼと接触させる請求項 105 に記載の方法。

【請求項 107】

4-アミノ-4-デオキシコリスミン酸シンターゼが p a p A を含む請求項 106 に記載の方法。

【請求項 108】

段階 (b) において4-アミノ-4-デオキシコリスミン酸をコリスミン酸ムターゼと接触させる請求項 105 に記載の方法。

【請求項 109】

コリスミン酸ムターゼが p a p B を含む請求項 108 に記載の方法。

【請求項 110】

段階 (c) において4-アミノ-4-デオキシプレフェン酸をプレフェン酸デヒドロゲナーゼと接触させる請求項 105 に記載の方法。

【請求項 111】

プレフェン酸デヒドロゲナーゼが p a p C を含む、請求項 110 に記載の方法。

【請求項 112】

段階 (d) においてp-アミノフェニルピルビン酸をアミノトランスフェラーゼと接触させる請求項 105 に記載の方法。

【請求項 113】

アミノトランスフェラーゼが大腸菌由来非特異的チロシンアミノトランスフェラーゼを含む請求項 112 に記載の方法。

【請求項 114】

アミノトランスフェラーゼが t y r B、a s p S 又は i l v E を含む請求項 116 に記載の方法。

【請求項 115】

段階 (a) ~ (d) を i n v i v o で実施する、請求項 105 に記載の方法。

【請求項 116】

p a p A、p a p B 及び p a p C を含むプラスミドで細胞を形質転換して大腸菌内でp-アミノフェニルアラニンを生産する方法であって、細胞がコリスミン酸とアミノトランスフェラーゼを含み、p a p A、p a p B 及び p a p C の発現の結果としてシンターゼ、ムターゼ及びデヒドロゲナーゼが生産され、シンターゼ、ムターゼ、デヒドロゲナーゼ及びアミノトランスフェラーゼがコリスミン酸からp-フェニルアラニンを生産する前記方法。

【請求項 117】

- (a) 細胞内の1種以上の炭素源から非天然アミノ酸を生産するための生合成経路系と、
- (b) 直交 t R N A (O - t R N A) と直交アミノアシル t R N A シンテターゼ (O - R S) を含む翻訳系を含む細胞であって、O - R S が O - t R N A を非天然アミノ酸で優先的にアミノアシル化し、O - t R N A がセクターコードンに応答して非天然アミノ酸をタンパク質に組込む前記細胞。

【請求項 118】

10

20

30

40

50

細胞が細菌細胞を含む請求項 1 1 7 に記載の細胞。

【請求項 1 1 9】

セクターコドンがナンセンスコドン、4 塩基コドン又はアンバーコドンを含む請求項 1 1 7 に記載の細胞。

【請求項 1 2 0】

ナンセンスコドンが T A G コドンを含む請求項 1 1 9 に記載の細胞。

【請求項 1 2 1】

生合成経路系が天然細胞量の非天然アミノ酸を生産する請求項 1 1 7 に記載の細胞。

【請求項 1 2 2】

生合成経路系がシンターゼ、ムターゼ、デヒドロゲナーゼ及びアミノトランスフェラーゼを含む請求項 1 1 7 に記載の細胞。 10

【請求項 1 2 3】

シンターゼが 4 - アミノ - 4 - デオキシコリスミン酸シンターゼを含む請求項 1 2 2 に記載の細胞。

【請求項 1 2 4】

ムターゼがコリスミン酸ムターゼを含む請求項 1 2 2 に記載の細胞。

【請求項 1 2 5】

デヒドロゲナーゼがプレフェン酸デヒドロゲナーゼを含む請求項 1 2 2 に記載の細胞。

【請求項 1 2 6】

シンターゼ、デヒドロゲナーゼ及びムターゼが *Streptomyces venezuelae* 又は *Streptomyces pristinaespiralis* に由来し、アミノトランスフェラーゼが大腸菌に由来する請求項 1 2 2 に記載の細胞。 20

【請求項 1 2 7】

非天然アミノ酸が p - アミノフェニルアラニン、d o p a 又は O - メチル - L - チロシンを含む請求項 1 1 7 に記載の細胞。

【請求項 1 2 8】

非天然アミノ酸がグリコシル化アミノ酸又はペグ化アミノ酸を含む請求項 1 1 7 に記載の細胞。

【請求項 1 2 9】

炭素源がコリスミン酸を含む請求項 1 1 7 に記載の細胞。 30

【請求項 1 3 0】

生合成経路系がコリスミン酸から p - アミノフェニルアラニンを生産する請求項 1 1 7 に記載の細胞。

【請求項 1 3 1】

生合成経路系が、

- (a) コリスミン酸を 4 - アミノ - 4 - デオキシコリスミン酸に酵素変換し、
- (b) 4 - アミノ - 4 - デオキシコリスミン酸を 4 - アミノ - 4 - デオキシプレフェン酸に酵素変換し、
- (c) 4 - アミノ - 4 - デオキシプレフェン酸を p - アミノフェニルピルビン酸に酵素変換し、
- (d) p - アミノフェニルピルビン酸を p - アミノフェニルアラニンに酵素変換する請求項 1 1 7 に記載の細胞。 40

【請求項 1 3 2】

細胞がタンパク質生合成に十分な量でありながら、天然アミノ酸の濃度を実質的に変えたり、細胞資源を枯渇させない量の非天然アミノ酸を生産する請求項 1 1 7 に記載の細胞。

【請求項 1 3 3】

非天然アミノ酸が p - アミノフェニルアラニン、L - d o p a 又は O - メチル - L - チロシンである請求項 1 3 2 に記載の細胞。

【請求項 1 3 4】

生合成経路系がプラスミドを含み、プラスミドがコリスミン酸シンターゼをコードするポ 50

リヌクレオチドと、コリスミン酸ムターゼをコードするポリヌクレオチドと、プレフェン酸デヒドロゲナーゼをコードするポリヌクレオチドを含む請求項 1 1 7 に記載の細胞。

【請求項 1 3 5】

細胞がプラスミドを含み、プラスミドが p a p A 遺伝子、p a p B 遺伝子及び p a p C 遺伝子を含む請求項 1 1 7 に記載の細胞。

【請求項 1 3 6】

プラスミドが更に l p p プロモーターと l a c プロモーターを含む請求項 1 3 5 に記載の細胞。

【請求項 1 3 7】

プラスミドが更に 1 個以上のリボソーム結合部位を含む請求項 1 3 5 に記載の細胞。 10

【請求項 1 3 8】

プラスミドが低コピー p S C 1 0 1 由来プラスミドである請求項 1 3 5 に記載の細胞。

【請求項 1 3 9】

少なくとも 2 1 種のアミノ酸を生産し、アミノ酸の 1 種以上を細胞内の 1 種以上のタンパク質に特異的に組込むための細胞であって、組込まれるアミノ酸の少なくとも 1 種が非天然アミノ酸を含む前記細胞。

【請求項 1 4 0】

細胞に組込まれた非天然アミノ酸により提供される効果を識別する方法であって、

(a) 大腸菌ゲノムに由来するランダム変異プラスミド (ランダム変異プラスミドの 1 個以上は非天然アミノ酸をタンパク質に組込む能力を細胞に付与する) を各々含む細胞のライブラリーを提供する段階と、 20

(b) 細胞ライブラリーをスクリーニングし、天然大腸菌細胞に比較して増殖が亢進した 1 個以上の細胞を識別し、非天然アミノ酸により付与される効果を識別する段階を含む前記方法。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

関連出願の相互参照

本出願は、2 0 0 1 年 4 月 1 9 日に提出された米国仮特許出願第 6 0 / 2 8 5 , 0 3 0 号、及び 2 0 0 2 年 2 月 6 日に提出された米国仮特許出願第 6 0 / 3 5 5 , 5 1 4 号の優先権および利益を主張し、それらの明細書の全体をここで援用する。 30

【0 0 0 2】

連邦政府の後援を受けた研究開発に関する声明

本発明は、海軍研究事務所からの補助金交付第 6 5 0 2 5 7 3 号及び国立衛生研究所からの補助金交付第 G M 6 2 1 5 9 号の下で米国政府の支援により為された。米国政府はこの発明に一定の権利を有する。

【0 0 0 3】

技術分野

本発明はタンパク質生合成の分野に関する。特に、本発明は非天然アミノ酸を含むタンパク質を生産するための組成物と方法の分野に関する。 40

【背景技術】

【0 0 0 4】

タンパク質は光合成からシグナル伝達や免疫応答に至るまでの複雑な生命現象のほぼ全てを行っている。これらの複雑な機能を理解し、制御するためには、タンパク質の構造と機能の関係の理解を深めることが必要である。

【0 0 0 5】

化合物の機能特性を変えるためにほぼ任意構造を変化させることができる小有機分子合成と異なり、タンパク質合成は 2 0 種の天然アミノ酸によりコードされる変化に限られる。細菌からヒトに至る全公知生物の遺伝コードは同一の 2 0 種の共通アミノ酸をコードする。これらのアミノ酸はタンパク質の翻訳後修飾 (例えばグリコシル化、リン酸化又は酸化 50

）、又は頻度は低いが、例えばセレノシステインの場合のようにアミノアシル化サプレッサー tRNA の酵素修飾により改変することができる。それにもかかわらず、これらの僅か 20 種の単純な構成単位から合成されるポリペプチドが複雑な生命現象の全てを行っている。

【0006】

部位特異的及びランダム突然変異誘発はタンパク質の特定アミノ酸を他の 19 種の共通アミノ酸の任意のもので置換でき、タンパク質の構造と機能の関係を解明するための重要なツールになっている。これらの方法により、安定性、触媒活性及び結合特異性等の特性を強化したタンパク質を作製することが可能になっている。しかし、タンパク質置換は大半が単純な官能基をもつ 20 種の共通アミノ酸に限られている Knowles, J. R. Tinker with enzymes: what are we learning? Science, 236 (4806) 1252-1258 (1987); 及び Zoller, M. J., Smith, M. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, Methods Enzymol, 154: 468-500 (1983) 参照。新規な生物学的、化学的又は物理的特性をもつ付加アミノ酸を含むように遺伝コードを拡張することにより、例えば天然タンパク質のように 20 種の共通アミノ酸に含まれるアミノ酸のみから構成されるタンパク質に比較してタンパク質の性質（例えば寸法、酸性度、求核性、水素結合、疎水性）を改変することができる。

10

【0007】

非天然アミノ酸をタンパク質に導入するために数種の戦略が利用されている。最初の実験は Lys、Cys 及び Tyr 等の反応性側鎖をアミノ酸に付加した（例えばリジンを N²-アセチル-L-リジンに変換）。化学合成も非天然アミノ酸を導入する簡単な方法であるが、日常的固相ペプチド合成は一般に 100 残基未満の小ペプチド又はタンパク質に限られている。ペプチドフラグメントの酵素ライゲーションと天然化学ライゲーションの最近の開発に伴い、より大きいタンパク質を生産することが可能であるが、方法の大規模化は容易ではない。例えば P. E. Dawson と S. B. H. Kent, Annu. Rev. Biochem., 69: 923 (2000) 参照。タンパク質生合成に有用な in vitro 抽出物に所望非天然アミノ酸で化学的にアシル化したサプレッサー tRNA を加える一般的な in vitro 生合成法を使用して 100 種を上回る非天然アミノ酸がほぼ任意寸法の各種タンパク質に部位特異的に組込まれている。例えば V. W. Cornish, D. Mendel and P. G. Schultz, Angew. Chem. Int. Ed. Engl., 1995, 34: 621 (1995); C. J. Norren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins, Science 244 182-188 (1989); 及び J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin, E. S. Dimala, Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide, J. Am. Chem. Soc. 111 8013-8014 (1989) 参照。タンパク質安定性、タンパク質フォールディング、酵素メカニズム及びシグナル伝達の研究のために多様な官能基がタンパク質に導入されている。これらの研究はタンパク質生合成機構が多様なアミノ酸側鎖を許容することを立証しているが、方法は技術的に困難であり、突然変異タンパク質収率は低い。

20

30

40

【0008】

50 年以上前に天然アミノ酸の多数の類似体は細菌の増殖を阻害することが発見された。これらのアミノ酸類似体の存在下で生産したタンパク質を分析すると、各種程度までその天然対応物を置換していることが分かった。例えば M. H. Richmond, Bacteriol. Rev., 26: 398 (1962) 参照。これは正しいアミノ酸とそのコ

50

グネイト tRNA の結合に関与する酵素であるアミノアシル-tRNA シンテターゼが類似体に対応する天然アミノ酸から厳密に区別できないためである。例えば、ノルロイシンはメチオニル-tRNA シンテターゼにより負荷され、p-フルオロフェニルアラニンはフェニルアラニン-tRNA シンテターゼにより負荷される。D. B. Cowie, G. N. Cohen, E. T. Bolton and H. DeRobinchon-Szulmajst, Biochim. Biophys. Acta, 1959, 34:39 (1959); 及び R. Munier と G. N. Cohen, Biochim. Biophys. Acta, 1959, 31:378 (1959) 参照。

【0009】

その後、種々の野生型シンテターゼを利用するために選択圧組込みと呼ばれる *in vivo* 方法が開発された。例えば N. Budisa, C. Minks, S. Alefeld, W. Wenger, F. M. Dong, L. Moroder and R. Huber, FASEB J., 13:41 (1999) 参照。この方法では特定天然アミノ酸を細胞に供給する該代謝経路を遮断する栄養要求株を使用し、ターゲット遺伝子の転写を抑制しながら制限濃度の天然アミノ酸を含む最少培地で培養する。定常増殖期が開始したら天然アミノ酸の供給を停止して非天然アミノ酸類似体に交換する。組換えタンパク質の発現を誘導すると、非天然類似体を含むタンパク質が蓄積する。例えば、このストラテジーを使用して o、m 及び p-フルオロフェニルアラニンがタンパク質に組込まれ、紫外線スペクトルで容易に識別可能な 2 つの特徴的な肩を示しており (例えば C. Minks, R. Huber, L. Moroder and N. Budisa, Anal. Biochem., 284:29 (2000) 参照)、また、トリフルオロメチオニンを使用してバクテリオファージ T4 リゾチームのメチオニンを置換し、キトオリゴ糖リガンドとのその相互作用が¹⁹F NMR により試験されており (例えば H. Duewel, E. Darb, V. Robinson and J. F. Honek, Biochemistry, 36:3404 (1997))、更にロイシンの代わりにトリフルオロロイシンを挿入してロイシンジッパータンパク質の熱及び化学安定性を高めている。例えば Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado and D. A. Tirrell, Angew. Chem. Int. Ed. Engl., 40:1494 (2001) 参照。更に、セレノメチオニンとテルロメチオニンを各種組換えタンパク質に組込んで X 線結晶解析で位相決定を容易にしている。例えば W. A. Hendrickson, J. R. Horton and D. M. Lemaster, EMBO J., 9:1665 (1990); J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odom, B. Dunlap, L. Leioda and M. Hatada, Nat. Struct. Biol., 1:283 (1994); N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann and R. Huber, Eur. J. Biochem., 230:788 (1995); 及び N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neufeind, L. Moroder and R. Huber, J. Mol. Biol., 270:616 (1997) 参照。アルケン又はアルキン官能基をもつメチオニン類似体も効率的に挿入され、化学的手段によりタンパク質の付加修飾が可能である。例えば J. C. M. van Hest と D. A. Tirrell, FEBS Lett., 424:68 (1998); J. C. M. van Hest, K. L. Kiick and D. A. Tirrell, J. Am. Chem. Soc., 122:1282 (2000); 及び K. L. Kiick と D. A. Tirrell, Tetrahedron, 56:9487 (2000) 参照。

【0010】

この方法の成功は一般にタンパク質翻訳の忠実度を確保するためには高い選択性を必要とするアミノアシル-tRNA シンテターゼによる非天然アミノ酸類似体の認識に依存している。従って、この方法により接近可能な化学官能基の範囲は限られている。例えば、チ

アプロリンは定量的にタンパク質に組込むことができるが、オキサプロリンやセレノプロリンはできない。N. Budisa, C. Minks, F. J. Medrano, J. Lutz, R. Huber and L. Moroder, Proc. Natl. Acad. Sci. U S A, 95: 455 (1998) 参照。この方法の範囲を拡大する1つの方法はアミノアシル-tRNAシンテターゼの基質特異性を緩和することであり、少数の例で成功している。例えば、大腸菌フェニルアラニル-tRNAシンテターゼ (PheRS) で Ala²⁹⁴ を Gly に置換すると、基質結合ポケットの寸法が増し、その結果、tRNA Phe は p-Cl-フェニルアラニン (p-Cl-Phe) によりアシル化される。M. Ibbá, P. Kast and H. Hennecke, Biochemistry, 33: 7107 (1994) 参照。この突然変異体 PheRS を大腸菌株に導入すると、フェニルアラニンの代わりに p-Cl-フェニルアラニン又は p-Br-フェニルアラニンを組込むことができる。例えば M. Ibbá と H. Hennecke, FEBS Lett., 364: 272 (1995); 及び N. Sharma, R. Furrer, P. Kast and D. A. Tirrell, FEBS Lett., 467: 37 (2000) 参照。同様に、大腸菌チロシル-tRNAシンテターゼのアミノ酸結合部位の近傍の点突然変異 Phe130Ser の結果、アザチロシンをチロシンよりも効率的に組込めることが示されている。F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll and S. Nishimura, J. Biol. Chem., 275: 40324 (2000) 参照。

10

20

【0011】

アミノアシル化の忠実性は非コグネイト中間体及び生成物の基質識別とブルーフリーディングの両レベルで維持される。従って、非天然アミノ酸をタンパク質に *in vivo* 組込む代替戦略はブルーフリーディングメカニズムをもつシンテターゼの改変である。これらのシンテターゼはコグネイト天然アミノ酸に構造的に類似するアミノ酸を識別することができず、従って活性化することができない。このエラーは別部位で修正され、tRNA から誤って負荷されたアミノ酸を脱アシル化し、タンパク質翻訳の忠実度を維持する。シンテターゼのブルーフリーディング活性が機能しなくなると、誤って活性化された構造類似体が編集機能から抜け落ちて組込まれる恐れがある。このアプローチはバリー-tRNAシンテターゼ (ValRS) で最近立証されている。V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel and P. Marlier, Science, 292: 501 (2001) 参照。ValRS が tRNA Val を Cys、Thr 又はアミノ酪酸 (Abu) で誤ってアミノアシル化すると、これらの非コグネイトアミノ酸は編集ドメインにより加水分解される。大腸菌染色体のランダム突然変異誘発後に、ValRS の編集部位に突然変異をもつ突然変異体大腸菌株が選択された。この編集欠損 ValRS は tRNA Val に誤って Cys を負荷する。Abu は Cys に立体的に似ている (Abu では Cys の -SH 基を -CH₃ で置換) ので、この突然変異体大腸菌株を Abu の存在下に増殖させると、突然変異体 ValRS は Abu もタンパク質に組込む。質量分析によると、天然タンパク質ではバリンの約 24% が各バリン位で Abu により置換されている。

30

40

【0012】

上記方法の少なくとも1つの重大な欠点は、タンパク質中の特定天然アミノ酸に対応する全部位を置換することである。また、天然及び非天然アミノ酸の組込み度も一定せず、細胞の内側でコグネイト天然アミノ酸を完全に枯渇させることは困難であるので、定量的置換が達成されることは稀である。類似体の多重部位組込みは毒性を生じるので、これらの戦略は生きた細胞では突然変異体タンパク質を試験し難いという欠点もある。最後に、ゲノム内の全部位で置換できなければならないという理由から、この方法は一般に共通アミノ酸の近似構造類似体にしか適用できない。

【0013】

50

固相合成及び半合成法により新規アミノ酸を含む多数の小タンパク質の合成も行われている。例えば以下の刊行物とその引用文献を参照されたい。Crick, F. J. C., Barrett, L. Brenner, S. Watts-Tobin, R. General nature of the genetic code for proteins. Nature, 1227-1232 (1961); Hofmann, K., Bohn, H. Studies on polypeptides. XXXVI. The effect of pyrazole-imidazole replacements on the S-protein activating potency of an S-peptide fragments, J. Am. Chem., 5914-5919 (1966); Kaiser, E. T. Synthetic approaches to biologically active peptides and proteins including enzymes, Acc Chem Res, 47-54 (1989); Nakatsuka, T., Sasaki, T., Kaiser, E. T. Peptide segment coupling catalyzed by the semisynthetic enzyme thiosubstilisins, J Am Chem Soc., 3808-3810 (1987); Schnolzer, M., Kent, S. B. H. Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease, Science, 221-225 (1992); Chaiken, I. M. Semisynthetic peptides and proteins, CRC Crit Rev Biochem, 255-301 (1981); Offord, R. E. Protein engineering by chemical means? Protein Eng., 151-157 (1987); 及び Jackson, D. Y., Burnier, J., Quan, C., Stanley, M., Tom, J., Wells, J. A. A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues, Science, 243 (1994)。

10

20

【0014】

補因子、スピンラベル及びオリゴヌクレオチド等の各種非天然側鎖をタンパク質に in vitro 導入するために化学修飾が使用されている。例えば Corey, D. R., Schultz, P. G. Generation of a hybrid sequence-specific single-stranded deoxyribonuclease, Science, 1401-1403 (1987); Kaiser, E. T., Lawrence D. S., Rokita, S. E. The chemical modification of enzymatic specificity, Rev Biochem, 565-595 (1985); Kaiser, E. T., Lawrence, D. S. Chemical mutation of enzyme active sites, Science, 505-511 (1984); Neet, K. E., Nanci A, Koshland, D. E. Properties of thiol-substilisins, J Biol. Chem., 6392-6401 (1968); Polgar, L. B., M. L. A new enzyme containing a synthetically formed active site. Thiol-substilisins. J. Am. Chem Soc., 3153-3154 (1966); 及び Pollack, S. J., Nakayama, G. Schultz, P. G. Introduction of nucleophiles and spectroscopic probes into antibody combining sites, Science, 1038-1040 (1988) 参照。

30

40

【0015】

あるいは、数種の生物学的プローブを in vitro 合成タンパク質に組込むために

50

化学修飾アミノアシル-tRNAを利用する生合成法が使用されている。以下の刊行物とその引用文献参照。Brunner, J. New Photolabeling and crosslinking methods, Annu. Rev. Biochem., 483-514 (1993); 及び Krieg, U. C., Walter, P., Hohnson, A. E. Photocrosslinking of the signal sequence of nascent preprolactin of the 54-kilodalton polypeptide of the signal recognition particle, Proc. Natl. Acad. Sci., 8604-8608 (1986)。

【0016】

所望アンバーナンセンス突然変異を含む遺伝子でプログラムされたタンパク質合成反応に化学的にアミノアシル化したサプレッサ-tRNAを加えることにより非天然アミノ酸をタンパク質に部位特異的に *in vitro* 組込むことができることは従来示されている。これらのアプローチを使用すると、特定アミノ酸の栄養要求性株を使用して20種の共通アミノ酸の多数のものを近似構造相同体（例えばフェニルアラニンをフルオロフェニルアラニン）で置換することができる。例えば Noren, C. J., Anthony-Cahill, M. C., Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins, Science, 244:182-188 (1989); M. W. Nowakら, Science 268:439-442 (1995); Bain, J. D., Glabe, C. G., Dix, T. A., Chamberlin, A. R., Diala, E. S. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide, J. Am. Chem. Soc., 111:8013-8014 (1989); N. Budisaら, FASEB J. 13:41-51 (1999); Ellman, J. A., Mendel, D., Anthony-Cahill, S., Noren, C. J., Schultz, P. G. Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. Methods in Enz., 301-336 (1992); 及び Mendel, D., Cornish, V. W. & Schultz, P. G. Site-Directed Mutagenesis with an Expanded Genetic Code, Annu. Rev. Biophys. Biomol. Struct. 24, 435-62 (1995) 参照。

【0017】

例えば、終止コドンUAGを認識するサプレッサ-tRNAが作製され、非天然アミノ酸で化学的にアミノアシル化された。タンパク質遺伝子の所期部位に終止コドンTAGを導入するために慣用部位特異的突然変異誘発が使用された。例えば Sayers, J. R., Schmidt, W. Eckstein, F. 5', 3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis, Nucleic Acids Res., 791-802 (1988) 参照。アシル化サプレッサ-tRNAと突然変異体遺伝子を *in vitro* 転写／翻訳系で組合せると、UAGコドンに応答して非天然アミノ酸が組込まれ、このアミノ酸を特定位置に含むタンパク質が得られた。[³H]-Pheを使用する実験と α -ヒドロキシ酸を使用する実験によると、所望アミノ酸のみがUAGコドンにより指定される位置に組込まれ、このアミノ酸はタンパク質の他のどの位置にも組込まれないことが判明した。例えば Norenら, 前出; 及び Ellman, J. A., Mendel, D., Schultz, P. G. Site-specific incorporation of novel backbone structures into proteins, Science, 197-200 (1992) 参照。

【0018】

一般に、これらの *in vitro* アプローチはアミノ酸の部位特異的組込みが困難であることや、アミノ酸が20種の共通アミノ酸の単純な誘導体でなければならないことや、大きいタンパク質又はペプチドフラグメントの合成に伴う固有の問題などの限界がある。

【0019】

非天然アミノ酸をタンパク質に組込むためにマイクロインジェクション法も使用されている。例えばM. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. G. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty and H. A. Lester, Science, 268:439 (1995); 及びD. A. Dougherty, Curr. Opin. Chem. Biol., 4:645 (2000) 参照。*in vitro* 作製した2種のRNA種即ち所期アミノ酸位置にUAG終止コドンをもつターゲットタンパク質をコードするmRNAと、所望非天然アミノ酸でアミノアシル化したアンバーサプレッサーtRNAをアフリカツメガエル卵母細胞に同時に注射している。こうすると、卵母細胞の翻訳機構はUAGにより指定される位置に非天然アミノ酸を挿入する。この方法により一体的膜タンパク質の*in vivo* 構造-機能研究が可能になったが、一般に*in vitro* 発現系には応用できない。例えば、蛍光アミノ酸をタキキニンニューロキニン-2受容体に組込んで蛍光共鳴エネルギー移動により距離を測定したり(例えばG. Turcatti, K. Nemeth, M. D. Edgerton, U. Meseth, F. Talabot, M. Peitsch, J. Knowles, H. Vogel and A. Chollet, J. Biol. Chem., 271:19991 (1996) 参照)、ピオチン化アミノ酸を組込んでイオンチャンネルで表面残基を同定したり(例えばJ. P. Gallivan, H. A. Lester and D. A. Dougherty, Chem. Biol., 4:739 (1997) 参照)、ケージドチロシン類似体を使用してイオンチャンネルの共焦点変化をリアルタイムでモニターしたり(例えばJ. C. Miller, S. K. Silverman, P. M. England, D. A. Dougherty and H. A. Lester, Neuron, 20:619 (1998) 参照)、 α ヒドロキシアミノ酸を使用してイオンチャンネルのバックボーンを改変し、その開閉メカニズムを探索する(例えばP. M. England, Y. Zhang, D. A. Dougherty and H. A. Lester, Cell, 96:89 (1999); 及びT. Lu, A. Y. Ting, J. Mainland, L. Y. Jan, P. G. Schultz and J. Yang, Nat. Neurosci., 4:239 (2001) 参照)等の方法が挙げられる。

【0020】

しかし、マイクロインジェクション法には、例えばサプレッサーtRNAを非天然アミノ酸で化学的に*in vitro* アミノアシル化しなければならないことや、アシル化tRNAが翻訳中に化学量論的試薬として消費され、再生できないなどの欠点がある。この欠点により、抑圧効率が悪く、タンパク質収率が低いので、電気生理学的測定等の高感度技術で突然変異体タンパク質を定量することが必要である。更に、この方法はマイクロインジェクションが可能な細胞にしか適用できない。

【0021】

非天然アミノ酸をタンパク質に直接*in vivo* 組込むことができるならば、突然変異体タンパク質の収率が上がり、技術的に容易であり、細胞や場合によっては生きた生物で突然変異体タンパク質を試験することが可能になり、これらの突然変異体タンパク質を治療処置に使用できる等の利点がある。各種寸法、酸性度、求核性、疎水性等をもつ非天然アミノ酸をタンパク質に組込むことができるならば、タンパク質の構造を合理的且つ体系的に操作し、タンパク質機能を探索すると共に新規特性をもつ新規タンパク質又は生物を創製する可能性が著しく広がる。しかし、タンパク質翻訳の忠実度を高くするために必要なtRNA-シンテターゼ相互作用の複雑な性質によりその方法は困難である。

【0022】

パラ-F-Pheを部位特異的に組込む試みの1つとして、酵母アンバーサプレッサー tRNA Phe CUA / フェニルアラニル-tRNA シンテターゼ対が p-F-Phe 耐性 Phe 栄養要求性大腸菌株で使用されている。例えば R. F u r t e r, P r o t e i n S c i., 7: 419 (1998) 参照。酵母 Phe RS は p-F-Phe に対して高い基質特異性をもたないので、突然変異誘発部位は 64~75% の p-F-Phe でしか翻訳されず、増殖培地に過剰の p-F-Phe を加えても Phe と L y s として残る。更に、Phe コドン位置に 7% の p-F-Phe が検出され、内在大腸菌 Phe RS は Phe 以外に p-F-Phe も組込むと考えられる。このアプローチは翻訳忠実度が劣り、例えば サプレッサー tRNA と Phe RS が真に直交ではないので、一般に他の非天然アミノ酸には適用できない。

10

【発明の開示】

【発明が解決しようとする課題】

【0023】

従って、細胞の生合成機構をより効率的且つ有効に改変できるように方法を改善することが必要である。以下の開示から自明の通り、本発明はこれらの必要及び他の必要に取り組むものである。

【課題を解決するための手段】

【0024】

本発明は非天然アミノ酸をタンパク質に組込むことができる翻訳系を作製及び使用するための各種方法と関連組成物を提供する。翻訳系により生産された非天然アミノ酸を含むタンパク質も本発明の特徴の1つである。本発明の翻訳系を使用すると、公知及び新規両者の非天然アミノ酸をタンパク質に組込むことができる。本発明は更に新規非天然アミノ酸、非天然アミノ酸を含む各種組成物（例えば非天然アミノ酸を含むタンパク質と細胞）、非天然アミノ酸を生産するための化学及び生化学的方法、並びに自律性 21 アミノ酸細胞の作製方法とこれを含む組成物も提供する。

20

【0025】

即ち、1側面において本発明は翻訳系を含む組成物を提供する。翻訳系は直交 tRNA (O-tRNA) と直交アミノアシル tRNA シンテターゼ (O-RS) を含む。一般に、O-RS は翻訳系の少なくとも1種の非天然アミノ酸で O-tRNA を優先的にアミノアシル化し、O-tRNA は少なくとも1種のセクターコドンを認識する。従って、翻訳系はコード化セクターコドンに応答して系で生産されるタンパク質に非天然アミノ酸を挿入する。

30

【0026】

典型的翻訳系としては細胞が挙げられ、例えば細菌細胞（例えば大腸菌）、始原菌細胞、真核細胞（例えば酵母細胞、哺乳動物細胞、植物細胞、昆虫細胞）等が挙げられる。あるいは、翻訳系は i n v i t r o 翻訳系を含み、例えば細胞抽出物を含む翻訳抽出物が挙げられる。

【0027】

O-tRNA の例としては、配列番号 1~3 及び／又はその相補的ポリヌクレオチド配列から構成される群から選択されるポリヌクレオチド配列を含む核酸が挙げられる。同様に、O-RS の例としては、配列番号 35~66 から構成される群から選択されるアミノ酸配列を含むポリペプチドと、配列番号 4~34 及びその相補的ポリヌクレオチド配列から構成される群から選択されるポリヌクレオチド配列を含む核酸によりコードされるポリペプチドから構成される群から選択されるポリペプチドが挙げられる。

40

【0028】

翻訳系により使用可能な非天然アミノ酸の例としては、チロシンアミノ酸の非天然類似体；グルタミンアミノ酸の非天然類似体；フェニルアラニンアミノ酸の非天然類似体；セリンアミノ酸の非天然類似体；スレオニンアミノ酸の非天然類似体；アルキル、アリール、アシル、アジド、シアノ、ハロ、ヒドラジン、ヒドラジド、ヒドロキシル、アルケニル、

50

アルキニル、エーテル、チオール、スルホニル、セレノ、エステル、チオ酸、硼酸、ボロン酸、ホスホ、ホスホノ、ホスフィン、複素環、エノン、イミン、アルデヒド、ヒドロキシルアミン、ケトもしくはアミノ置換アミノ酸又はその任意組合せ；光架橋基をもつアミノ酸；スピンラベル付きアミノ酸；蛍光アミノ酸：新規官能基をもつアミノ酸；別の分子と共有又は非共有的に相互作用するアミノ酸；金属結合性アミノ酸；金属含有アミノ酸；放射性アミノ酸；フォトケージド及び／又は光異性化可能なアミノ酸；ビオチン又はビオチン類似体含有アミノ酸；グリコシル化又は糖鎖修飾アミノ酸；ケト含有アミノ酸；ポリエチレングリコール又はポリエーテルを含むアミノ酸；重原子置換アミノ酸；化学分解性又は光分解性アミノ酸；延長側鎖をもつアミノ酸；毒性基を含むアミノ酸；糖鎖置換アミノ酸（例えば糖鎖置換セリン等）；炭素結合糖含有アミノ酸；レドックス活性アミノ酸； α -ヒドロキシ含有酸；アミノチオ酸含有アミノ酸； α , α ジ置換アミノ酸； β アミノ酸；及びプロリン以外の環状アミノ酸が挙げられる。

10

【0029】

例えば、非天然アミノ酸としてはO-メチル-L-チロシン、L-3-(2-ナフチル)アラニン、3-メチルフェニルアラニン、O-4-アリル-L-チロシン、4-プロピル-L-チロシン、トリ-O-アセチル-GlcNAc β -セリン、L-Dopa、フッ素化フェニルアラニン、イソプロピル-L-フェニルアラニン、p-アジド-L-フェニルアラニン、p-アシル-L-フェニルアラニン、p-ベンゾイル-L-フェニルアラニン、L-ホスホセリン、ホスホノセリン、ホスホノチロシン、p-ヨードフェニルアラニン、p-ブロモフェニルアラニン、p-アミノ-L-フェニルアラニン、及びイソプロピル-L-フェニルアラニンが挙げられる。1態様では、少なくとも1種の非天然アミノ酸はO-メチル-L-チロシンである。1特定態様では、少なくとも1種の非天然アミノ酸はL-3-(2-ナフチル)アラニンである。別の特定態様では、少なくとも1種の非天然アミノ酸はアミノ、イソプロピル、又はO-アリル含有フェニルアラニン類似体である。

20

【0030】

本発明では各種セクターコドンの任意のものを使用することができ、例えばナンセンスコドン、レアコドン又は4塩基（以上）コドン等が挙げられる。例えば、1態様では少なくとも1種のセクターコドンはアンバーコドンである。

【0031】

本明細書には各種翻訳系の例を示し、例えばmtRNA_{CUA}^{Tyr}と突然変異体TyrRS(LWJ16)を含む大腸菌細胞が挙げられ、この場合、突然変異体TyrRS(LWJ16)は細胞内でmtRNA_{CUA}^{Tyr}をO-メチル-L-チロシンで優先的にアミノアシル化し、細胞はmtRNA_{CUA}^{Tyr}を使用してアンバーコドンを認識する。別の例ではmtRNA_{CUA}^{Tyr}とSS12-TyrRSを含む大腸菌細胞が挙げられ、この場合、SS12-TyrRSは細胞内でmtRNA_{CUA}^{Tyr}をL-3-(2-ナフチル)アラニンで優先的にアミノアシル化し、細胞はmtRNA_{CUA}^{Tyr}を使用してアンバーコドンを認識する。

30

【0032】

本発明の翻訳系は有用量の大量の非天然アミノ酸を含むタンパク質を合成することができる。例えば、細胞抽出物と緩衝液と医薬的に許容可能な賦形剤及び／又はその他を含む組成物中に少なくとも約10、50、100 μ g/L以上の濃度で少なくとも1種の非天然アミノ酸を含むタンパク質を生産することができる。

40

【0033】

本発明の別の側面は入手可能な任意タンパク質と相同であるが、1種以上の非天然アミノ酸相同体を含むタンパク質を生産できるようにする。例えば、1種以上の非天然アミノ酸を含み、1種以上の治療用タンパク質に相同の治療用タンパク質を生産することができる。例えば、1側面では、タンパク質はサイトカイン、増殖因子、増殖因子受容体、インターフェロン、インターロイキン、炎症性分子、腫瘍遺伝子産物、ペプチドホルモン、シグナル伝達分子、ステロイドホルモン受容体、転写アクチベーター、転写サプレッサー、エ

50

リスロポエチン (EPO)、インスリン、ヒト成長ホルモン、上皮好中球活性化ペプチド
 -78、GRO α /MGSA、GRO β 、GRO γ 、MIP-1 α 、MIP-1 δ 、MCP-1、肝細胞増殖因子、インスリン様増殖因子、白血病阻害因子、オンコスタチンM、
 PD-ECSF、PDGF、プレイオトロピン、SCF、cキットリガンド、VEGEF、G-CSF、IL-1、IL-2、IL-8、IGF-I、IGF-II、FGF (繊維芽細胞増殖因子)、PDGF、TNF、TGF- α 、TGF- β 、EGF (表皮増殖因子)、KGF (ケラチノサイト増殖因子)、SCF/cキット、CD40L/CD40、
 VLA-4/VCAM-1、ICAM-1/LFA-1、ヒアルリン/CD44、Mos、Ras、Raf、Met、p53、Tat、Fos、Myc、Jun、Myb、Rel
 、エストロゲン受容体、プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、LDL受容体、及び/又はコルチコステロン受容体等の治療用又は他のタンパク
 質に相同である。別の態様では、タンパク質は α 1アンチトリプシン、アンジオスタチン
 、抗血友病因子、抗体、アポリポタンパク質、アポタンパク質、心房性ナトリウム利尿因
 子、心房性ナトリウム利尿ポリペプチド、心房性ペプチド、C-X-Cケモカイン、T3
 9765、NAP-2、ENA-78、Gro-a、Gro-b、Gro-c、IP-1
 0、GCP-2、NAP-4、SDF-1、PF-4、MIG、カルシトニン、cキット
 リガンド、サイトカイン、CCケモカイン、単球化学誘引タンパク質-1、単球化学誘引
 タンパク質-2、単球化学誘引タンパク質-3、単球炎症性タンパク質-1 α 、単球炎症
 性タンパク質-1 β 、RANTES、I309、R83915、R91733、HCC1
 、T58847、D31065、T64262、CD40、CD40リガンド、Cキット
 リガンド、コラーゲン、コロニー刺激因子 (CSF)、補体因子5a、補体阻害剤、補体
 受容体1、サイトカイン、上皮好中球活性化ペプチド-78、GRO α /MGSA、GR
 O β 、GRO γ 、MIP-1 α 、MIP-1 δ 、MCP-1、表皮増殖因子 (EGF)、
 上皮好中球活性化ペプチド、エリスロポエチン (EPO)、表皮剥離毒素、IX因子、V
 II因子、VIII因子、X因子、繊維芽細胞増殖因子 (FGF)、フィブリノーゲン、
 フィブネクチン、G-CSF、GM-CSF、グルコセレブロシダーゼ、ゴナドトロピ
 ン、増殖因子、増殖因子受容体、ヘッジホッグタンパク質、ヘモグロビン、肝細胞増殖因
 子 (HGF)、ヒルジン、ヒト血清アルブミン、ICAM-1、ICAM-1受容体、L
 FA-1、LFA-1受容体、インスリン、インスリン様増殖因子 (IGF)、IGF-I、
 IGF-II、インターフェロン、IFN- α 、IFN- β 、IFN- γ 、インター
 ロイキン、IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、
 IL-8、IL-9、IL-10、IL-11、IL-12、ケラチノサイト増殖因子 (KGF)、
 ラクトフェリン、白血病阻害因子、ルシフェラーゼ、ニールチュリン、好中
 球阻害因子 (NIF)、オンコスタチンM、骨形成タンパク質、腫瘍遺伝子産物、副甲状
 腺ホルモン、PD-ECSF、PDGF、ペプチドホルモン、ヒト増殖ホルモン、プレ
 イオトロピン、プロテインA、プロテインG、発熱外毒素A、B又はC、リラキシン、レニ
 ン、SCF、可溶性補体受容体I、可溶性ICAM1、可溶性インターロイキン受容体
 、可溶性TNF受容体、ソマトメジン、ソマトスタチン、ソマトトロピン、ストレプトキ
 ナーゼ、スーパー抗原、ブドウ球菌エンテロトキシン、SEA、SEB、SEC1、SE
 C2、SEC3、SED、SEE、ステロイドホルモン受容体、スーパーオキシドジスム
 ターゼ、毒素性ショック症候群毒素、チモシン α 1、組織プラスミノゲンアクチベーター、
 腫瘍増殖因子 (TGF)、TGF- α 、TGF- β 、腫瘍壊死因子、腫瘍壊死因子 α
 、腫瘍壊死因子 β 、腫瘍壊死因子受容体 (TNFR)、VLA-4タンパク質、VCAM
 -1タンパク質、血管内皮増殖因子 (VEGEF)、ウロキナーゼ、Mos、Ras、R
 af、Met、p53、Tat、Fos、Myc、Jun、Myb、Rel、エストロゲ
 ン受容体、プロゲステロン受容体、テストステロン受容体、アルドステロン受容体、LD
 L受容体、及び/又はコルチコステロン受容体等の治療用又は他のタンパク質に相同であ
 る。1側面では、本発明の組成物は非天然アミノ酸を含むタンパク質と医薬的に許容可能
 な賦形剤を含み、例えば上記タンパク質の任意のものと医薬的に許容可能な賦形剤を含む
 。

【0034】

ポリペプチドとの相同度は例えばデフォルトパラメーターに設定した例えばBLASTI
N又はBLASTPを使用して配列アラインメントを実施することにより推定することが
できる。例えば、1態様では、タンパク質は公知治療用タンパク質（例えばGeneban
k又は他の入手可能なデータベースに登録されたタンパク質）に少なくとも約50%、
少なくとも約75%、少なくとも約80%、少なくとも約90%又は少なくとも約95%
一致する。例えば、1好適態様では、治療用タンパク質はエリスロポエチン（EPO）で
ある。

【0035】

所期タンパク質は1、2、3、4、5、6、7、8、9、10、11、12、13、14 10
、15又は16種以上の非天然アミノ酸を含むことができる。非天然アミノ酸は同一でも
異なってもよく、例えば1、2、3、4、5、6、7、8、9、10、11、12、
13、14、15又は16種以上の異なる非天然アミノ酸を含むタンパク質に1、2、3
、4、5、6、7、8、9、10、11、12、13、14、15又は16個以上の異なる
部位が存在することができる。例えば、1態様では、タンパク質はDHFRであり、少
なくとも1種の非天然アミノ酸はO-メチル-L-チロシンとL-3-（2-ナフチル）
アラニンから構成される群から選択される。

【0036】

本発明は少なくとも1種のタンパク質が少なくとも1種の非天然アミノ酸を含むように少
なくとも1種のタンパク質を翻訳系で生産する方法も提供する。本方法では、少なくとも 20
1種のセレクトターコドンを含む少なくとも1種の核酸を翻訳系に加え、前記核酸は前記少
なくとも1種のタンパク質をコードする。翻訳系で機能し、少なくとも1種のセレクトター
コドンを認識する直交tRNA（O-tRNA）と、翻訳系内でO-tRNAを少なくと
も1種の非天然アミノ酸で優先的にアミノアシル化する直交アミノアシルtRNAシンテ
ターゼ（O-RS）も翻訳系に加える。少なくとも1種の非天然アミノ酸も翻訳系に加え
、少なくとも1種の非天然アミノ酸を含む少なくとも1種のタンパク質を翻訳系で生産す
る。

【0037】

本方法では組成物の上記全構造特徴、例えば翻訳系の種類（例えば細胞、細胞抽出物等）
、翻訳系で生産されるタンパク質の種類（例えばEPO相同体及び本明細書に記載する他 30
のタンパク質）、特定突然変異体タンパク質、特定非天然アミノ酸等を具体化することが
できる。

【0038】

1側面では、生産される非天然アミノ酸を含むタンパク質を細胞依存的にプロセッシング及
び改変する。こうして、安定的にフォールディング、グリコシル化又は他の方法で改変さ
れたタンパク質を細胞により生産することができる。

【0039】

非天然アミノ酸は場合により外部から翻訳系に加える。あるいは、例えば翻訳系が細胞で
ある場合には、非天然アミノ酸を翻訳系により生合成することができる。

【0040】

1特定態様では、本発明は少なくとも1種のO-メチル-L-チロシンを含む少なくとも
1種のタンパク質を大腸菌細胞で生産する方法を提供する。本方法はアンバーコドンを含
み、前記少なくとも1種のタンパク質をコードする少なくとも1種の核酸を翻訳系に加え
る段階と、mtRNA_{CUA}^{Tyr}（但し、mtRNA_{CUA}^{Tyr}は細胞で機能し、mtRNA_{CUA}^{Tyr}はアンバーコドンを認識する）を翻訳系に加える段階と、突然変異体TyrRS
（LWJ16）（但し、突然変異体TyrRS（LWJ16）は細胞内でmtRNA_{CUA}^{Tyr}をO-メチル-L-チロシンでアミノアシル化する）を翻訳系に加える段階と、O-
メチル-L-チロシンを細胞に加え、O-メチル-L-チロシンを含む少なくとも1種の
タンパク質を細胞内で生産する段階を含む。

【0041】

40

50

別の態様では、本発明は少なくとも1種のL-3-(2-ナフチル)アラニンを含む少なくとも1種のタンパク質を大腸菌細胞で生産する方法を提供する。本態様では、本方法はアンバーコドンを含み、前記少なくとも1種のタンパク質をコードする少なくとも1種の核酸を翻訳系に加える段階と、 $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ (但し、 $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ は細胞で機能し、 $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ はアンバーコドンを認識する) を細胞に加える段階と、 SS12-TyrRS (但し、 SS12-TyrRS は細胞内で $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ をL-3-(2-ナフチル)アラニンでアミノアシル化する) を細胞に加える段階と、L-3-(2-ナフチル)アラニンを細胞に加え、L-3-(2-ナフチル)アラニンを含む少なくとも1種のタンパク質を細胞内で生産する段階を含む。

【0042】

別の側面では、本発明は例えばメタ置換フェニルアラニン類似体 (例えば3-アセチルフェニルアラニン及び3-メトキシフェニルアラニン)、チロシン類似体 (例えば4-アリルチロシン)、グリコシル化アミノ酸等の非天然アミノ酸を提供する。

【0043】

非天然アミノ酸を含む各種組成物 (例えば本発明の非天然アミノ酸を含むタンパク質や細胞) も提供する。例えば、(例えば共有結合した) 非天然アミノ酸と直交tRNAを含む組成物を提供する。(例えば水素結合した) 非天然アミノ酸と直交アミノアシルtRNAシントターゼを含む組成物も提供する。

【0044】

1側面では、本発明はアミノ酸の合成方法を提供する。例えば、4-アリル-L-チロシンを合成するには一般に例えば水素化ナトリウムとDMFの存在下に保護チロシンを臭化アリルと反応させ、脱保護して4-アリル-L-チロシンを得る。一般に、NBoc又はFmoc保護チロシンを使用し、例えば塩酸とジオキサンの存在下に例えば酸脱保護する。最終生成物の場合により例えばエタノール又はジクロロメタンで抽出する。

【0045】

メタ置換フェニルアラニン類似体を合成するには一般にアセトアミドマロン酸ジエチルとメタ置換臭化ベンジルを縮合させる。その後、縮合物を一般に加水分解してメタ置換フェニルアラニン類似体 (例えばケト、アセチル又はメトキシ置換フェニルアラニン、例えば3-メトキシフェニルアラニン又は3-アセチルフェニルアラニン) を得る。所望メタ置換臭化ベンジルは、場合によりN-ブロモスクシンイミド (NBS) を3-メチルアセトフェノンと反応させて臭素化物を生成し、臭素化物をヘキサン溶液中で結晶化させることにより合成される。結晶化させると、一臭化物と二臭化物の混合物ではなく一臭化物の生成物が得られる。

【0046】

別の側面では、本発明は非天然アミノ酸の生合成方法を提供する。例えば、場合により例えばN-アセチルガラクトサミニダーゼ、トランスグリコシラーゼ又はセリングリコシヒドロラーゼの遺伝子を含むプラスミドで細胞を形質転換させることによりグリコシル化アミノ酸を*in vivo*合成する。その後、細胞は例えば細胞資源から所望グリコシル化アミノ酸を生産する。別の例では、例えばコリスミン酸を4-アミノ-4-デオキシコリスミン酸に酵素変換し、4-アミノ-4-デオキシコリスミン酸を4-アミノ-4-デオキシプレフェン酸に酵素変換し、4-アミノ-4-デオキシプレフェン酸をp-アミノフェニルピルビン酸に酵素変換し、p-アミノフェニルピルビン酸をp-アミノフェニルアラニンに酵素変換することによりp-アミノフェニルアラニンを*in vivo*合成する。酵素変換は一般に夫々4-アミノ-4-デオキシコリスミン酸シントターゼ (例えばPapA)、コリスミン酸ムターゼ (例えばPapB) 及びプレフェン酸デヒドロゲナーゼ (例えばPapC) を使用して実施する。最終段階は一般にp-アミノフェニルピルビン酸をアミノトランスフェラーゼ (例えば大腸菌由来非特異的チロシンアミノトランスフェラーゼ) と接触させることにより実施する。本発明で使用するアミノトランスフェラーゼとしてはtyrB、aspS又はilvEが挙げられるが、これらに限定されない。一般に上記段階は例えば合成に使用する酵素をコードする遺伝子を含むプラスミドで細胞を形

10

20

30

40

50

質転換させることにより *in vivo* で実施する。

【0047】

別の側面では、本発明は大腸菌で *p*-アミノフェニルアラニンを生産する方法を提供する。本方法は一般にコリスミン酸とアミノトランスフェラーゼを含む細胞を *papA*、*papB* 及び *papC* を含むプラスミドで形質転換する。*papA*、*papB* 及び *papC* の発現の結果としてシンターゼ、ムターゼ及びデヒドロゲナーゼが生産され、これらの酵素がアミノトランスフェラーゼと共にコリスミン酸から *p*-フェニルアラニンを生産する。

【0048】

別の側面では、本発明は自律 21（以上）アミノ酸細胞を提供する。細胞（例えば細菌細胞）は一般に細胞内の 1 種以上の炭素源（例えばコリスミン酸）から非天然アミノ酸（例えば *p*-アミノフェニルアラニン）を生産するための生合成経路系と、直交 *tRNA*（*O*-*tRNA*）と直交アミノアシル *tRNA* シンテターゼ（*O*-*RS*）を含む翻訳系を含む。*O*-*RS* は *O*-*tRNA* を非天然アミノ酸で優先的にアミノアシル化し、*O*-*tRNA* はセクターコドン（例えば *TAG* 等のナンセンスコドン、4 塩基コドン又はアンバーコドン）に応答して非天然アミノ酸をタンパク質に組込む。細胞は 1 種以上（例えば 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15 又は 16 種以上）の非天然アミノ酸を含むことができ、場合により 1 種以上の直交 *tRNA*（例えば非天然アミノ酸をタンパク質に 1 個ずつ部位特異的に組込むには非天然アミノ酸 1 個につき 1 個、又は非天然アミノ酸組込みの特異性を調節するにはそれ以上もしくはそれ以下）及び／又は 1 種以上の直交アミノアシル *tRNA* シンテターゼ（*O*-*RS*）（例えば直交 *tRNA* 1 個につき 1 個、又は非天然アミノ酸組込みの特異性を調節するにはそれ以上もしくはそれ以下）も含むことができる。

【0049】

所定態様では、生合成経路系は天然細胞量の非天然アミノ酸を生産し、例えば細胞はタンパク質生合成に十分な量でありながら、非天然アミノ酸の生産において天然アミノ酸の濃度を実質的に変えたり、細胞資源を実質的に枯渇させない量の非天然アミノ酸を生産する。

【0050】

1 類の態様では、上記のようにコリスミン酸から *p*-アミノフェニルアラニンを生産するように自律細胞を遺伝子組換えする。本態様では、上記のような所望酵素（例えば *Streptomyces venezuelae* 又は *Streptomyces pristinaespiralis* に由来するシンターゼ、デヒドロゲナーゼ及びムターゼと大腸菌に由来するアミノトランスフェラーゼ）を生産するように細胞を遺伝子組換えする。例えば、本発明の細胞は、場合により *papA*、*papB* 及び *papC* と更に *lpp* プロモーターと *lac* プロモーターを含むプラスミド（例えば低コピー *pSC101* 由来プラスミド）で形質転換する。所定態様では、プラスミドは更に 1 個以上のリボソーム結合部位を含む。

【0051】

本発明の細胞により場合により生産される他の非天然アミノ酸としては、*dopa*、*O*-メチルー *L*-チロシン、グリコシル化アミノ酸、ペグ化アミノ酸、本明細書に記載する他の非天然アミノ酸等が挙げられるが、これらに限定されない。

【0052】

別の関連側面では、本発明は少なくとも 21 種以上のアミノ酸を生産し、アミノ酸の 1 種以上を細胞内の 1 種以上のタンパク質に特異的に組込むための 1 種以上の系を含む細胞も提供し、組込まれるアミノ酸の少なくとも 1 種は非天然アミノ酸を含む。

【0053】

別の側面では、本発明は細胞の 1 種以上のタンパク質に組込まれた非天然アミノ酸により提供される効果を識別する方法を提供する。本方法は一般に例えば大腸菌ゲノムに由来するランダム変異プラスミドを各々含む細胞のライブラリーを提供する段階を含む。ランダム変異プラスミドの 1 個以上は一般に非天然アミノ酸をタンパク質に組込み能力を細胞に

10

20

30

40

50

付与する。次に細胞ライブラリーをスクリーニングし、天然大腸菌細胞に比較して増殖が亢進した細胞を識別し、非天然アミノ酸により付与される効果を識別する。所定態様では、第2のスクリーニングを使用し、識別された任意効果が非天然アミノ酸に起因することを更に確認する。

【0054】

本発明の付加特徴はキットである。例えば、キットは1種以上の上記翻訳系（例えば細胞、21以上のアミノ酸の細胞等）と、1種以上の非天然アミノ酸と、例えば適当なパッケージング材料と、キットの成分を保持するための容器と、本発明の方法を実施するための説明書及び／又は他の成分を含むことができる。同様に、例えばキットの成分を保持するための容器と、本発明の方法を実施するための説明書及び／又は他の成分を含むキットの形態で翻訳系の生産物（例えば非天然アミノ酸を含むEPO類似体等のタンパク質）を提供することもできる。 10

【0055】

図面の簡単な説明

図1は、Ty r R Sの活性部位におけるアミノ酸残基の立体図（P. B r i c k, T. N. B h a t, D. M. B l o w, J. M o l. B i o l. 208, 83-98 (1988) を改変）。B. s t e a r o t h e r m o p h i l u s Ty r R Sに由来する残基も記入した。

【0056】

図2は、各種条件下における野生型（w t）及び突然変異体両者の大腸菌D H F Rタンパク質の蓄積を示す。発現条件は各レーンの上部に示す。左側のレーンは分子量マーカである。図2 Aは精製D H F Rの銀染色S D S-P A G Eゲルである。図2 Bは図2 Aのゲルのウェスタンブロットである。 20

【0057】

図3は、D H F RのN H 2末端ペプチドM I Y* M I A A L A V D Rのタンデム質量スペクトル。O-メチル-L-チロシン残基（Y*）を含むペプチドの部分配列Y* M I A A L A V D Rは注釈付きb又はyイオン系列から読み取ることができる。

【0058】

図4は、各種条件下における野生型（w t）及び突然変異体両者のマウスD H F Rタンパク質の蓄積を示す。発現条件は各レーンの上部に示す。左側のレーンは分子量マーカである。図4 Aは精製D H F Rの銀染色S D S-P A G Eゲルである。図4 Bは図4 Aのゲルのウェスタンブロットである。 30

【0059】

図5は、トリプシン消化ペプチドL L P E X* T G V L S E V Q E E K（式中、X*はL-3-（2-ナフチル）アラニンを表す）のタンデム質量スペクトル。配列は注釈付きb又はyイオン系列から読み取ることができるが、b 7とy 1 3は観測されない。二重負荷y 1 4に帰属する基準ピーク8 2 1. 7（100%）は分かり易くするために短縮している。

【0060】

図6 A～Dは、増幅性蛍光レポーター系の特徴を示す。図6 Aはプラスミドp R E Pである。T 7 R N Aポリメラーゼ転写はa r aプロモーターにより制御され、タンパク質発現は遺伝子内の各種位置のアンバーコドンの抑圧に依存する。G F P u v発現はT 7 R N Aポリメラーゼにより制御される。プラスミドp R E Pは直交シンテターゼ／t R N A対を発現するC o l E 1プラスミドと使用するのに適合可能である。図6 Bはp R E P（1～12）内のT 7 R N Aポリメラーゼ遺伝子構築物の組成と蛍光増加を示す。構築物番号を各構築物の左側に示す。蛍光増加は各構築物の右側に示し、p R E P（1～12）とp Q又はp Q Dを含む細胞の蛍光をフルオロメトリーにより測定し、細胞濃度補正比として計算した。遺伝子内のアンバー突然変異の位置を各構築物の上部に示す。図6 Cはp R E P（10）とp Q D（上段）又はp Q（下段）を含む細胞のサイトメトリー分析を示す。図6 Dはp R E P（10）を含み、各種大腸菌サプレッサーt R N Aを発現する細胞の蛍光 40 50

分析を示す。「なし」は細胞がサプレッサー tRNA を含まないことを示す。

【0061】

図7A～Cは、M. jannaschii TyrRSの進化を誘導するための多目的レポータープラスミド系の成分を示す。図7Aはプラスミド pREP/YC-JYCUAを示す。プラスミド pREP/YC-JYCUAはプラスミド pBK及び変異体と使用するのに適合可能である。図7BはM. jannaschii TyrRSの進化のターゲットとして使用した非天然アミノ酸の構造を示す。図7Cはプラスミド pREP/YC-JYCUAを使用してアミノアシル-tRNAシンテターゼを進化させるためのストラテジーを示す。蛍光及び非蛍光細胞を夫々黒とグレーで示す。

【0062】

図8A～Dは、成功した各進化実験からの優性シンテターゼ変異体の活性を示す。図8AはpREP/YC-JYCUAと指定シンテターゼ変異体を含む細胞を対応する非天然アミノ酸の存在下(+)又は不在下(-)で増殖させ、長波長紫外線を照射した場合の写真である。図8BはpREP/YC-JYCUAと指定シンテターゼ変異体を含む細胞を対応する非天然アミノ酸の存在下(左)又は不在下(右)で増殖させた場合の蛍光分析を示す。図8CはpREP/YC-JYCUAと指定シンテターゼ変異体を含む細胞を対応する非天然アミノ酸の存在下又は不在下で増殖させた場合のCm IC₅₀分析を示す表である。図8DはpBAD/JYAMB-4TAGと指定シンテターゼ変異体を含む細胞を対応する非天然アミノ酸の存在下(+)又は不在下(-)で増殖させた場合のタンパク質発現分析を示す。

【0063】

図9は、ネガティブFACSスクリーニング[OAY-RS(1, 3, 5)]又はネガティブバルナーゼ選択[OAY-RS(B)]を使用して得られたOAS-RS変異体の活性比較を示す。pREP/YC-JYCUAと指定シンテターゼ変異体を含む細胞を対応する非天然アミノ酸の存在下(立体棒、左)又は不在下(立体棒、右)で増殖させ、蛍光分析した。非天然アミノ酸の不在下に対して存在下で増殖させた細胞の蛍光の細胞濃度補正比として蛍光増加(背後棒)計算した。

【0064】

図10は、m-MeO-Pheとm-アセチル-Pheを組込んだDHFRの発現を示すウェスタンブロットのオートラジオグラフ。

【0065】

図11は、フルオレセインヒドラジドで標識したタンパク質の蛍光スペクトルを示す。

【0066】

図12は、非天然アミノ酸パラアジドフェニルアラニン及びパラベンゾイルフェニルアラニンを示す。

【0067】

図13は、アリル置換フェニルアラニンの合成の為の化学スキームを示す。

【0068】

図14は、メタ置換フェニルアラニンの合成のための化学スキームを示す。

【0069】

図15は、p-アミノフェニルアラニンの生合成を示す。Aはp-アミノフェニルアラニンの生合成に使用するプラスミドを示し、Bは例えばAのプラスミドを使用してコリスミン酸からp-アミノフェニルアラニンを生産するための生合成スキームを示す。

【0070】

図16は、各種非天然アミノ酸を示す。

【0071】

図17は、各種非天然アミノ酸を示す。

【0072】

図18は、各種非天然アミノ酸を示す。

【0073】

10

20

30

40

50

図 19 は、i n v i v o 抑圧によりタンパク質に組込む付加的天然及び非天然アミノ酸を示す。

【0074】

図 20 は、d o p a の生産のための生合成スキームを示す。

【0075】

図 21 は、21 種のアミノ酸を特異的に組込むことができる結果として細胞に得られる進化的効果を測定するための方法を示す。

【0076】

図 22 は、非天然アミノ酸の部位特異的組込み方法を示す。

【0077】

図 23 は、各種グルタミン類似体の合成を示す。

【0078】

図 24 は、 γ 置換グルタミン類似体の合成を示す。

【0079】

図 25 は、環状グルタミン誘導体の合成を示す。

【0080】

図 26 は、各種チロシン類似体を示す。

【0081】

図 27 は、チロシン類似体の生産のための生合成スキームを示す。

【0082】

図 28 は、グリコシル化アミノ酸を生産するための生合成スキームを示す。

【0083】

図 29 は、例えば細胞取込み試験で使用する各種非天然アミノ酸を示す。以上の任意又は全図面は実際に模式図である。

【発明を実施するための最良の形態】

【0084】

概要

本発明は直交 t R N A / アミノアシル t R N A シンテターゼ (O - t R N A / O - R S) 対を使用して遺伝的にコードされる付加アミノ酸を受容するように細胞のタンパク質生合成機構を補強するための組成物と方法を提供する。本明細書に記載する組成物と方法は非天然アミノ酸と共に使用し、例えば多様な側鎖の任意のものを使用して新規立体、化学又は構造特性をタンパク質に提供することができる。本発明は原核 (例えば真正細菌、始原菌) 細胞と真核 (例えば酵母、哺乳動物、植物又は昆虫) 細胞の両者に適用可能である。これらの組成物と方法はセレクトターコドン (例えば終止コドン、4 塩基コドン等) を介して非天然アミノ酸を部位特異的に組込むために有用である。本発明は本発明の組成物を使用して生産されるか又は本発明の方法により生産された非天然アミノ酸を含むタンパク質も提供する。非天然アミノ酸を直接生きた細胞でタンパク質に導入できるので、タンパク質及び細胞機能の研究の新規ツールを提供すると共に、例えば治療に有用な特性を増進したタンパク質を作製することが可能になる。

【0085】

定義

相同 : タンパク質及び / 又はタンパク質配列は共通の祖先タンパク質又はタンパク質配列から天然又は人工的に誘導される場合に「相同」である。同様に、核酸及び / 又は核酸配列は共通の祖先核酸又は核酸配列から天然又は人工的に誘導される場合に相同である。例えば、入手可能な任意突然変異誘発法により任意天然核酸を改変し、1 個以上のセレクトターコドンを加えることができる。この突然変異誘発した核酸は発現されると、1 種以上の非天然アミノ酸を含むポリペプチドをコードする。突然変異法は当然のことながら更に 1 個以上の標準コドンを変異させ、得られる突然変異体タンパク質の 1 個以上の標準アミノ酸も変異させることができる。相同性は一般に 2 種以上の核酸又はタンパク質 (又はその配列) 間の配列類似度から推定される。相同性の判定に有用な配列間類似度の厳密な百分

10

20

30

40

50

率は該当核酸及びタンパク質により異なるが、通常は25%程度の低い配列類似度を使用して相同性を判定する。もっと高いレベルの配列類似度、例えば30%、40%、50%、60%、70%、80%、90%、95%又は99%以上を使用して相同性を判定することもできる。配列類似度百分率の決定方法（例えばデフォルトパラメーターを使用するBLASTP及びBLASTN）は本明細書に記載し、一般に入手可能である。

【0086】

直交：本明細書において「直交」なる用語は所期系（例えば翻訳系、例えば細胞）により低効率で使われる分子（例えば直交tRNA（O-tRNA）及び／又は直交アミノアシルtRNAシンテターゼ（O-RS））を意味する。直交とは直交tRNA及び／又は直交RSが所期翻訳系で機能できないか又は機能効率が低く、例えば20%未満、10%未満、5%未満、又は例えば1%未満であることを意味する。例えば、直交tRNAが所期翻訳系で所期翻訳系の任意内在RSをアミノアシル化する効率は内在RSによる内在tRNAのアミノアシル化に比較して低下するか又はゼロである。別の例では、直交RSが所期翻訳系で任意内在tRNAをアミノアシル化する効率は内在RSによる内在tRNAのアミノアシル化に比較して低下するか又はゼロである。

10

【0087】

優先的にアミノアシル化する：「優先的にアミノアシル化する」なる用語はO-RSが非天然アミノ酸でO-tRNAをアミノアシル化する効率が天然tRNA又はO-tRNAを作製するために使用する出発材料に比較した場合に例えば約70%、約75%、約85%、約90%、約95%、又は約99%以上であることを意味する。その後、非天然アミノ酸は所与セクターコドンで約75%以上、所与セクターコドンで約80%以上、所与セクターコドンで約90%以上、所与セクターコドンで約95%以上、又は所与セクターコドンで99%以上の高い忠実度で成長中のポリペプチド鎖に組込まれる。

20

【0088】

セクターコドン：「セクターコドン」なる用語は翻訳プロセスでO-tRNAにより認識されるが、内在tRNAには認識されないコドンを意味する。O-tRNAアンチコドンループはmRNA上のセクターコドンを認識し、そのアミノ酸（例えば非天然アミノ酸）をポリペプチド内のこの部位に組込む。セクターコドンとしては例えばナンセンスコドン、例えば終止コドン（例えばアンバー、オーカー及びオパールコドン）、4塩基以上のコドン、天然又は非天然塩基対から誘導されるコドン等を挙げることができる。所与系ではセクターコドンは更に天然3塩基コドンの1種を含むことができ、内在系は前記天然3塩基コドンを使用せず、このような系としては例えば天然3塩基コドンを認識するtRNAを欠失する系や、天然3塩基コドンがレアコドンである系が挙げられる。

30

【0089】

サプレッサーtRNA：サプレッサーtRNAは所与翻訳系でメッセンジャーRNA（mRNA）の読取りを改変するtRNAである。サプレッサーtRNAは例えば終止コドン、4塩基コドン又はレアコドンを介して読み取ることができる。

【0090】

翻訳系：「翻訳系」なる用語は成長中のポリペプチド鎖（タンパク質）に天然アミノ酸を組込むために必要な成分を意味する。翻訳系の成分としては例えばリボソーム、tRNA、シンテターゼ、mRNA等を挙げることができる。本発明の成分はin vivoでもin vitroでも翻訳系に加えることができる。翻訳系は原核細胞（例えば大腸菌）でも真核細胞（例えば酵母、哺乳動物、植物又は昆虫）でもよい。

40

【0091】

非天然アミノ酸：本明細書において「非天然アミノ酸」なる用語は20種の天然アミノ酸の1種又はセレノシステイン以外の任意アミノ酸、改変アミノ酸及び／又はアミノ酸類似体を意味する。

【0092】

本項又は本明細書の以下の記載において特に定義しない限り、本明細書で使用する全技術及び科学用語は本発明が属する分野の当業者に一般に理解されている通りの意味をもつ。

50

【0093】

詳細な説明

タンパク質は光合成や視覚からシグナル伝達や免疫応答までのほぼ全生物プロセスの岐路に位置する。これらの複雑な機能は規定一次配列に配置された比較的単純な20個の構成単位から構成されるポリアミド系ポリマーから得られる。

【0094】

本発明は非天然アミノ酸をタンパク質に部位特異的に直接 *in vivo* 組込むのに使用する方法と組成物を含む。重要な点は、非天然アミノ酸を20種の共通アミノ酸の1種に置換するのでなく遺伝子レパートリーに加えるという点である。本発明は例えば (i) 任意タンパク質の任意所望位置に1種以上の非天然アミノ酸を部位特異的又はランダムに挿入することができ、(ii) 原核細胞と真核細胞のいずれにも適用可能であり、(iii) 大量の精製突然変異体タンパク質の作製に加えて突然変異体タンパク質の *in vivo* 試験が可能であり、(iv) 多様な非天然アミノ酸の任意のものをタンパク質に *in vivo* 組込むように応用できる。本発明は非天然アミノ酸の *in vivo* 部位特異的組込みに有用な組成物と方法を提供する。具体的には、本発明は直交 tRNA (O-tRNA) と直交アミノアシル tRNA シンテターゼ (O-RS) と非天然アミノ酸を含む翻訳系 (例えば細胞) を提供し、O-RS は O-tRNA を非天然アミノ酸でアミノアシル化し、細胞は成長中のポリペプチド鎖に非天然アミノ酸を組込むために成分を使用する。

【0095】

本発明は更に本発明の翻訳系を使用して非天然アミノ酸を部位特異的に *in vivo* 組込むための方法を提供する。本発明は本発明の方法により生産されたタンパク質も提供する。請求するタンパク質は非天然アミノ酸を含む。

【0096】

本発明の組成物と方法は、直交 tRNA (O-tRNA) アミノアシル tRNA シンテターゼ (O-RS) 対を使用する。O-tRNA が O-RS により非天然アミノ酸で優先的にアミノアシル化される性質をもつ限り、多様な対を使用することができる。更に、直交対は所期翻訳系で機能し、例えば翻訳系は非天然アミノ酸でアミノアシル化された O-tRNA を使用して非天然アミノ酸をポリペプチド鎖に組込む。組込みは部位特異的に行われ、例えば O-tRNA はタンパク質をコードする mRNA でセクターコドン (例えば終止コドン) を認識する。

【0097】

1 態様において、O-tRNA は *Methanococcus jannaschii* 細胞に由来する Tyr-tRNA から誘導される。1 好適態様では、O-tRNA は本明細書で $mtRNA_{CUA}^{Tyr}$ と呼ぶものである。別の態様では、O-tRNA は配列番号 1 ~ 3 を含む群から選択されるポリヌクレオチド配列又はその相補的ポリヌクレオチド配列を含む核酸を含む。

【0098】

本発明の所定態様において、O-RS は *Methanococcus jannaschii* 細胞に由来する TyrRS から誘導される。1 好適態様では、O-RS は本明細書で突然変異体 TyrRS (LWJ16) 又は SS12-TyrRS と呼ぶものである。別の態様では、O-RS は、配列番号 35 ~ 66 から構成される群から選択されるアミノ酸配列を含むポリペプチドと、配列番号 4 ~ 34 から構成される群から選択されるポリヌクレオチド配列又はその相補的ポリヌクレオチド配列を含む核酸によりコードされるポリペプチドから構成される群から選択されるポリペプチドを含む。

【0099】

1 好適態様では、本発明は $mtRNA_{CUA}^{Tyr}$ と突然変異体 TyrRS (LWJ16) を含む大腸菌細胞に関し、突然変異体 TyrRS (LWJ16) は細胞内で $mtRNA_{CUA}^{Tyr}$ を O-メチル-L-チロシンで優先的にアミノアシル化し、細胞は $mtRNA_{CUA}^{Tyr}$ を使用してアンバーコドンを認識する。

【0100】

別の好適態様では、本発明は $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ と SS12-TyrRS を含む大腸菌細胞に関し、 SS12-TyrRS は細胞内で $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ を L-3-(2-ナフチル) アラニンで優先的にアミノアシル化し、細胞は $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ を使用してアンバーコドンを確認する。

【0101】

具体的な O-tRNA 及び O-RS 分子の配列は実施例に記載する。

【0102】

直交 tRNA と直交アミノアシル tRNA シンテターゼの対

直交対は O-tRNA (例えばサプレッサー tRNA 、フレームシフト tRNA 等) と O-RS から構成される。 O-tRNA は上述のように内在シンテターゼによりアシル化されず、セクターコドンでデコードすることができる。 O-RS は例えば拡張アンチコドンループで O-tRNA を認識し、 O-tRNA を非天然アミノ酸で優先的にアミノアシル化する。多重直交 tRNA / シンテターゼ対の開発により、各種コドンを使用して多重非天然アミノ酸を同時に組込むことが可能になる。

【0103】

O-tRNA と O-RS は天然に存在するものでもよいし、起源と宿主名で記載されている各種生物に由来する天然 tRNA 及び／又は RS を突然変異させることにより誘導してもよい。各種態様において、 O-tRNA と O-RS は少なくとも1種の生物に由来する。別の態様では、 O-tRNA は第1の生物に由来する天然に存在するか又は天然に存在するものを突然変異させた tRNA から誘導し、 O-RS は第2の生物に由来する天然に存在するか又は天然に存在するものを突然変異させた RS から誘導する。

【0104】

本発明の組成物と方法で使用する O-tRNA 、 O-RS 及び対を得るための方法 (誘導、突然変異、スクリーニング) は本願と同時に出願した米国特許出願番号第 _____ 号 (発明の名称 "Methods and Compositions for the production of orthogonal tRNA-tRNA synthetase pairs") (代理人整理番号 54-000130US, Schultz ら) にも記載されており、その開示内容全体を本明細書に組込む。

【0105】

これらの方法は直交 tRNA / RS 対を作製するために試みられた他のストラテジーについて背景技術の項に記載した問題を解決する。具体的には、これらの方法は (a) 第1の生物に由来する少なくとも1種の tRNA から誘導される tRNA のライブラリーを作製する段階と、(b) 第1の生物に由来する RS の不在下に第2の生物に由来するアミノアシル- tRNA シンテターゼ (RS) によりアミノアシル化される tRNA のライブラリーをネガティブ選択することにより tRNA のプールを提供する段階と、(c) 導入した直交 RS (O-RS) によりアミノアシル化される個体を tRNA のプールから選択することにより少なくとも1種の組換え O-tRNA を提供する段階を含む。少なくとも1種の組換え O-tRNA はセクターコドンで認識し、第2の生物に由来する RS により効率的に認識されず、 O-RS により優先的にアミノアシル化される。本方法は更に、(d) 第3の生物に由来する少なくとも1種のアミノアシル- tRNA シンテターゼ (RS) から誘導される突然変異体 RS のライブラリーを作製する段階と、(e) 非天然アミノ酸と天然アミノ酸の存在下に少なくとも1種の組換え O-tRNA を優先的にアミノアシル化するメンバーを RS のライブラリーから選択することにより活性 RS のプールを提供する段階と、(f) 非天然アミノ酸の不在下に少なくとも1種の組換え O-tRNA を優先的にアミノアシル化する活性 RS のプールをネガティブ選択することにより少なくとも1種の特異的 O-tRNA / O-RS 対を提供する段階を含み、但し、少なくとも1種の特異的 O-tRNA / O-RS 対は非天然アミノ酸に特異的な少なくとも1種の組換え O-RS と少なくとも1種の組換え O-tRNA を含む。

【0106】

直交対を作製する第1のストラテジーは突然変異体ライブラリーを作製してこのライブラ

リーからO-tRNA又はO-RSをスクリーニング及び／又は選択する。

【0107】

直交tRNA／シンテターゼ対を作製する第2のストラテジーは異種tRNA／シンテターゼ対（例えば別の起源の生物からの対）を宿主細胞に導入する。異種シンテターゼ候補の性質としては、例えば宿主細胞tRNAに負荷しない等が挙げられ、異種tRNA候補の性質としては、例えば宿主細胞シンテターゼによりアシル化されない等が挙げられる。更に、異種tRNAから誘導される異種tRNAは全宿主細胞シンテターゼに直交性である。

【0108】

本明細書及び米国特許出願番号第_____号（発明の名称"Methods and Compositions for the production of orthogonal tRNA-tRNA synthetase pairs"）（代理人整理番号54-000130US, Schultzら）に記載されている方法を使用し、上記所望対及び対の成分を進化させ、所望特性、例えばO-tRNAを非天然アミノ酸で優先的にアミノアシル化できるという特性をもつ直交tRNA／シンテターゼ対を作製する。 10

【0109】

本明細書では非天然アミノ酸をタンパク質に組込むためのストラテジーについて記載するが、同様にセレクトターコードンに応答して天然アミノ酸を組込むストラテジーを開発し、突然変異誘発の付加基盤を提供することも自明である。即ち、本明細書全般に記載する非天然アミノ酸の負荷と同様にセレクトターコードンを認識する直交tRNAに天然アミノ酸を負荷するようにシンテターゼを改変することができる。 20

【0110】

直交アミノアシルtRNAシンテターゼ（O-RS）の生産

O-RSの生産方法は野生型シンテターゼの骨格から突然変異体シンテターゼのプールを作製した後、20種の共通アミノ酸に比較して非天然アミノ酸に対する特異性に基づいて突然変異RSを選択することを基本とする。このようなシンテターゼを単離するために、本発明の選択方法は（i）1回目の選択からの所望シンテターゼの活性を低くすることができ、集団を小さくできるので高感度であり、（ii）各回の選択で選択ストリンジェンシーを変えることが望ましいので「調節可能」であり、（iii）汎用性であるため、種々の非天然アミノ酸に使用できる。 30

【0111】

直交アミノアシルtRNAシンテターゼを作製する方法は例えばシンテターゼの活性部位、シンテターゼの編集メカニズム部位、シンテターゼの種々のドメインを組み合わせることで種々の部位等でシンテターゼを突然変異させ、選択プロセスを適用する。ポジティブ選択後にネガティブ選択を行う併用に基づくストラテジーを使用する。ポジティブ選択では、ポジティブマーカの非必須位置に導入したセレクトターコードンを抑圧し、ポジティブ選択圧下に細胞を生存させる。従って、天然及び非天然アミノ酸両者の存在下で生存細胞は直交サプレッサーtRNAに天然又は非天然アミノ酸を負荷する活性シンテターゼをコードする。ネガティブ選択では、ネガティブマーカの非必須位置に導入したセレクトターコードンを抑圧し、天然アミノ酸特異性をもつシンテターゼを除去する。ネガティブ選択とポジティブ選択の生存細胞は直交サプレッサーtRNAを非天然アミノ酸のみでアミノアシル化（負荷）するシンテターゼをコードする。その後、これらのシンテターゼを例えばDNAシャフリング又は他の再帰的突然変異誘発法により更に突然変異誘発することができる。 40

【0112】

突然変異体RSのライブラリーは、当分野で公知の種々の突然変異誘発技術を使用して作製することができる。例えば、突然変異体RSは部位特異的突然変異、ランダム点突然変異、相同組換え、キメラ構築等により作製することができる。

【0113】

ポジティブ選択段階は例えば少なくとも1種のセレクトターコードン（例えばアンバーコードン 50

）を含むポジティブ選択マーカー（例えば抗生物質耐性遺伝子等）と突然変異体 R S のライブラリーを複数の細胞に導入し、選択物質の存在下に複数の細胞を増殖させ、ポジティブ選択マーカー中で少なくとも 1 種のセクターコドンを抑圧することにより選択物質の存在下に生存する細胞を選択し、活性突然変異体 R S のプールを含むポジティブ選択細胞のサブセットを提供することができる。場合により、選択物質濃度を変動してもよい。

【0114】

ネガティブ選択段階は例えば少なくとも 1 種のセクターコドンを含む抗生物質耐性遺伝子（例えばクロラムフェニコールアセチルトランスフェラーゼ（CAT）遺伝子）であるネガティブ選択マーカーをポジティブ選択からの活性突然変異体 R S のプールと共に第 2 の生物の複数の細胞に導入し、非天然アミノ酸と選択物質を加えた第 1 の培地で生存するが、非天然アミノ酸と選択物質を加えない第 2 の培地では生存することができない細胞を選択し、少なくとも 1 種の組換え O-R S をもつ生存細胞を提供することができる。場合により、選択物質濃度を変動してもよい。

10

【0115】

ポジティブ選択はセクターコドン（例えばアンバー終止コドン）を含むポジティブ選択マーカー（例えばクロラムフェニコールアセチルトランスフェラーゼ（CAT）遺伝子）の抑圧に基づいて行うことができるので、クロラムフェニコールをポジティブ選択圧として適用することができる。更に、非天然アミノ酸の存在下と不在下で本明細書に記載するようなポジティブマーカーとネガティブマーカーのいずれにも CAT 遺伝子を使用することができる。場合により、セクターコドンを含む CAT 遺伝子をポジティブ選択に使用し、少なくとも 1 種以上のセクターコドンを含むネガティブ選択マーカー（例えばバルナーゼ遺伝子等の毒性マーカー）をネガティブ選択に使用する。

20

【0116】

ポジティブ選択は細胞をアンピシリン耐性にする β -ラクタマーゼ遺伝子の非必須位置におけるセクターコドンを抑圧に基づいて行うこともでき、ネガティブマーカーとしてリボヌクレアーゼバルナーゼを使用するネガティブ選択を使用する。ペリプラズムに分泌される β -ラクタマーゼに対して、CAT は細胞質に局在し、更にアンピシリンは殺菌性であるが、クロラムフェニコールは静菌性である。

【0117】

組換え O-R S を更に突然変異させ、選択することができる。1 態様では、少なくとも 1 種の組換え直交アミノアシル-tRNA シンテターゼ（O-R S）の生産方法は更に（d）少なくとも 1 種の組換え O-R S を単離する段階と、（e）少なくとも 1 種の組換え O-R S から誘導される第 2 組の突然変異 O-R S を作製する段階と、（f）O-tRNA を優先的にアミノアシル化することができる突然変異 O-R S が得られるまで段階（b）及び（c）を反復する段階を含むことができる。場合により、段階（d）～（f）を例えば少なくとも約 2 回反復する。1 側面では、第 2 組の突然変異 O-R S は突然変異誘発（例えばランダム突然変異誘発、部位特異的突然変異誘発、組換え又はその組合せ）により作製することができる。

30

【0118】

直交 tRNA（O-tRNA）の生産

組換え直交 tRNA（O-tRNA）の生産方法は米国特許出願番号第 _____ 号（発明の名称 "Methods and Compositions for the production of orthogonal tRNA-tRNA synthetase pairs"）（代理人整理番号 54-000130US, Schultz ら）に記載されている。

40

【0119】

組換え O-tRNA の生産方法は（a）第 1 の生物に由来する少なくとも 1 種の tRNA（例えばサプレッサー tRNA）から誘導される突然変異体 tRNA のライブラリーを作製する段階と、（b）第 1 の生物に由来する R S の不在下に第 2 の生物に由来するアミノアシル-tRNA シンテターゼ（R S）によりアミノアシル化される突然変異体 tRNA

50

のライブラリーをネガティブ選択することにより突然変異体 tRNA のプールを提供する段階と、(c) 導入した直交 RS (O-RS) によりアミノアシル化されるメンバーを突然変異体 tRNA のプールから選択することにより少なくとも 1 種の組換え O-tRNA を提供する段階を含み、但し、少なくとも 1 種の組換え O-tRNA はセクターコドン を認識し、第 2 の生物に由来する RS により効率的に認識されず、O-RS により優先的にアミノアシル化される。1 態様では、組換え O-tRNA は直交性が改善されている。

【0120】

例えば、所望 RS に対するその親和性を維持しながら tRNA の直交性を改善するために、本方法は夫々コグネイトシンテターゼの不在下と存在下の突然変異体 サプレッサー tRNA によるネガティブ選択とポジティブ選択を組合せる。ネガティブ選択ではセクターコドンをマーカー遺伝子 (例えばバルナーゼ遺伝子等の毒性遺伝子) の非必須位置に導入する。例えば *Methanococcus jannaschii* に由来する突然変異 tRNA ライブラリーのメンバーが内在宿主 (例えば大腸菌) シンテターゼによりアミノアシル化される (即ち宿主、例えば大腸菌シンテターゼに非直交性である) 場合には、セクターコドン (例えばアンバーコドン) が抑圧され、生産された毒性遺伝子産物は細胞死に至る。直交 tRNA 又は非機能的 tRNA をもつ細胞は生存する。その後、生存細胞をポジティブ選択し、セクターコドン (例えばアンバーコドン) をポジティブマーカー遺伝子 (例えば β -ラクタマーゼ遺伝子等の薬剤耐性遺伝子) に挿入する。これらの細胞はコグネイト RS をもつ発現ベクターも含む。これらの細胞を選択物質 (例えばアンピシリン) の存在下に増殖させる。その後、同時発現したコグネイトシンテターゼによりアミノアシル化され、このセクターコドンに応答してアミノ酸を挿入することができる tRNA を選択する。非機能的 tRNA 又は所期シンテターゼにより認識することができない tRNA を含む細胞は抗生物質に感受性である。従って、(i) 内在宿主 (例えば大腸菌) シンテターゼの基質でなく、(ii) 所期シンテターゼによりアミノアシル化することができ、(iii) 翻訳に機能的な tRNA はどちらの選択でも生存する。

【0121】

突然変異 tRNA のライブラリーを構築する。tRNA の所望ループ (例えばアンチコドンループ (D アーム、V ループ、T Ψ C アーム) 又はループ組合せ又は全ループ) の特定位置 (例えば非保存位置、保存位置)、ランダム位置又は両者の組合せに突然変異を導入することができる。tRNA のキメラライブラリーも本発明に含まれる。場合により各種生物 (例えば真正細菌又は始原菌等の微生物) に由来する tRNA シンテターゼのライブラリー (例えば天然ダイバーシティを含むライブラリー) (例えば Short らの米国特許第 6, 238, 884 号、Schallenberg らの米国特許第 5, 756, 316 号、Petersen らの米国特許第 5, 783, 431 号、Thompson らの米国特許第 5, 824, 485 号、Short らの米国特許第 5, 958, 672 号参照) を構築して直交対をスクリーニングしてもよい。

【0122】

例えば、アミノアシルー tRNA シンテターゼによりアミノアシル化される突然変異体 tRNA のライブラリーをネガティブ選択するには、少なくとも 1 種以上のセクターコドンを含む毒性マーカー遺伝子と突然変異体 tRNA のライブラリーを第 2 の生物に由来する複数の細胞に導入し、少なくとも 1 種の直交 tRNA 又は非機能的 tRNA を含む突然変異体 tRNA のプールを含む生存細胞を選択することができる。例えば、毒性マーカー遺伝子はリボヌクレアーゼバルナーゼ遺伝子であり、リボヌクレアーゼバルナーゼ遺伝子は少なくとも 1 個のアンバーコドンを含む。場合により、リボヌクレアーゼバルナーゼ遺伝子に 2 個以上のアンバーコドンを加えてもよい。生存細胞は例えば比較細胞密度アッセイを使用することにより選択することができる。

【0123】

別の例では、導入した直交 RS (O-RS) によりアミノアシル化されるメンバーを突然変異体 tRNA のプールから選択するには、例えばセクターコドンの少なくとも 1 種を含む薬剤耐性遺伝子 (例えば β -ラクタマーゼ遺伝子) を含むポジティブ選択マーカー遺

伝子（例えば少なくとも1種のアンバー終止コドンを含む β -ラクタマーゼ遺伝子）とO-RSと突然変異体tRNAのプールを第2の生物に由来する複数の細胞に導入し、選択物質（例えば抗生物質）の存在下に増殖した生存細胞を選択することにより、O-RSによりアミノアシル化され、少なくとも1種のセレクトターコドンに応答してポジティブマーカ遺伝子によりコードされる翻訳産物にアミノ酸を挿入する少なくとも1種の組換えtRNAをもつ細胞プールを提供することができる。別の態様では、選択物質濃度を変動させる。本方法により生産された組換えO-tRNAも本発明に含まれる。

【0124】

上記方法における選択段階（例えばポジティブ選択段階、ネガティブ選択段階又はポジティブ選択段階とネガティブ選択段階の両者）のストリンジェンシーは場合により選択ストリンジェンシーの変動を含む。例えば、バルナーゼは極毒性タンパク質であるので、バルナーゼ遺伝子に導入するセレクトターコドンの数を変えることによりネガティブセレクトターのストリンジェンシーを制御することができる。本発明の1側面では、初期回に所望活性を低くできるのでストリンジェンシーを変動できる。即ち、初期回には低ストリンジェンシー選択基準を適用し、後期回の選択ではよりストリンジェントな基準を適用する。

【0125】

本発明では例えばO-RS、O-tRNA及びO-tRNA/O-RS対を作製するために他の型の選択も使用できる。例えば、ポジティブ選択段階、ネガティブ選択段階又はポジティブ選択段階とネガティブ選択段階の両者はレポーターを使用することができ、レポーターは蛍光活性化細胞ソーティング（FACS）により検出される。例えば、ポジティブ選択マーカ（例えばセレクトターコドン（例えばアンバー終止コドン）を含むクロラムフェニコールアセチルトランスフェラーゼ（CAT）遺伝子）を使用してまずポジティブ選択を行った後にネガティブ選択スクリーニングを行い、ネガティブマーカ（例えばT7 RNAポリメラーゼ遺伝子）内の位置でセレクトターコドン（例えば2個以上）を抑圧できないものを選択することができる。1態様では、ポジティブ選択マーカとネガティブ選択マーカを同一ベクター（例えばプラスミド）に配置することができる。ネガティブマーカの発現はレポーター（例えばグリーン蛍光タンパク質（GFP））の発現を誘導する。選択とスクリーニングのストリンジェンシーは変動させることができ、例えばレポーターを蛍光させるために必要な光強度を変動させることができる。別の態様では、FACによりスクリーニングされるポジティブ選択マーカとしてレポーターを使用してポジティブ選択を行った後にネガティブ選択スクリーニングを行い、ネガティブマーカ（例えばバルナーゼ遺伝子）内の位置でセレクトターコドン（例えば2個以上）を抑圧できないものを選択することができる。

【0126】

場合により、レポーターを細胞表面、ファージディスプレイ等に提示する。細胞表面提示（例えばOmpA細胞表面提示系）は大腸菌細胞表面における特定エピトープ（例えば例えば外膜ポリリンOmpAに融合したポリオウイルスCペプチド）の発現に依存する。エピトープは翻訳中にタンパク質メッセージ中のセレクトターコドンが抑圧されるときにしか細胞表面に提示されない。従って、提示されるペプチドはライブラリー内の突然変異体アミノアシル-tRNAシンテターゼの1種により認識されるアミノ酸を含み、特定非天然アミノ酸を含むペプチドに対する抗体で対応するシンテターゼ遺伝子を含む細胞を分離することができる。OmpA細胞表面提示系はファージディスプレイの代用としてGeorgiouらにより開発され、改良された。Francisco, J. A., Campbell, R., Iverson, B. L. & Georgiou, G. Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. Proc Natl Acad Sci U S A. 90:10444-8 (1993) 参照。

【0127】

選択段階は *in vitro* で実施することもできる。その後、選択した成分（例えばシンテターゼ及び／又は *tRNA*）を非天然アミノ酸の *in vivo* 組込みに使用する細胞に導入することができる。

【0128】

資源及び宿主生物

直交 *tRNA-RS* 対は例えば少なくとも第1の資源生物又は同一でも異なっているてもよい少なくとも2種の資源生物に由来し、各種宿主生物（例えば第2の生物）で使うことができる。本発明の方法の第1の生物と第2の生物は同一でも異なっているてもよい。1態様では第1の生物は例えば *Methanococcus jannaschii*、*Methanobacterium thermoautotrophicum*、*Halobacterium*、*Escherichia coli*、*A. fulgidus*、*Halobacterium*、*P. furiosus*、*P. horikoshii*、*A. pernix*、*T. thermophilus* 等の原核生物である。あるいは、第1の生物は例えば植物（例えば単子葉植物又は双子葉植物等の複雑な植物）、藻類、原生動物、真菌類（例えば酵母等）、動物（例えば哺乳動物、昆虫、節足動物等）等の真核生物でもよい。別の態様では、第2の生物は *Methanococcus jannaschii*、*Methanobacterium thermoautotrophicum*、*Halobacterium*、*Escherichia coli*、*A. fulgidus*、*Halobacterium*、*P. furiosus*、*P. horikoshii*、*A. pernix*、*T. thermophilus* 等の原核生物である。あるいは、第2の生物は例えば植物、真菌類、動物等の真核生物でもよい。

10

20

【0129】

上述のように、対の各成分は同一生物に由来するものでも異なる生物に由来するものでもよい。例えば、*tRNA* は例えば始原菌（例えば *Methanococcus jannaschii* 及び *Halobacterium NURC-1*）又は真正細菌（例えば大腸菌）等の原核生物に由来し、シンテターゼは同一又は異なる原核生物（例えば *Methanococcus jannaschii*、*Archaeoglobus fulgidus*、*Methanobacterium thermoautotrophicum*、*P. furiosus*、*P. horikoshii*、*A. pernix*、*T. thermophilus*、*Halobacterium*、*Escherichia coli* 等）に由来するものとすることができる。例えば植物（例えば単子葉植物又は双子葉植物等の複雑な植物）、藻類、原生動物、真菌類（例えば酵母等）、動物（例えば哺乳動物、昆虫、節足動物等）等の真核源も使用できる。

30

【0130】

セクターコードン

本発明のセクターコードンは、タンパク質生合成機構の遺伝コードンを拡張する。例えば、セクターコードンとしては例えばユニーク3塩基コードン、ナンセンスコードン、（例えばアンバーコードン又はオパールコードン等の終止コードン）、非天然コードン、少なくとも4塩基のコードン等が挙げられる。例えば1個以上、2個以上、3個以上等の多数のセクターコードンを所望遺伝子に導入することができる。

40

【0131】

64種の遺伝コードンは20種のアミノ酸と3種の終止コードンをコードする。翻訳終結には1個の終止コードンしか必要ないので、他の2個は主に非タンパク生産アミノ酸をコードするために使用することができる。アンバー終止コードンである *UAG* は *in vitro* 生合成系とアフリカツメガエル卵母細胞で使って非天然アミノ酸の組込みに成功している。3種の終止コードンのうちで *UAG* は大腸菌での使用頻度が最も低い終止コードンである。大腸菌株には *UAG* を認識して天然アミノ酸を挿入する天然サプレッサー *tRNA* を含むものもある。更に、これらのアンバーサプレッサー *tRNA* は慣用タンパク質突然変異誘発でも使われている。

【0132】

50

1 態様では、本方法は非天然アミノ酸の *in vivo* 組込みに終止コドンであるセクターコドンを使用する。例えば、終止コドン（例えば UAG）を認識する O-tRNA を作製し、O-RS により所望非天然アミノ酸でアミノアシル化する。この O-tRNA は天然アミノアシル-tRNA シンテターゼにより認識されない。慣用部位特異的突然変異誘発を使用して終止コドン（例えば TAG）をタンパク質遺伝子内の所期部位に導入することができる。例えば Sayers, J. R., Schmidt, W. Eckstein, F. 5', 3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res.*, 791-802 (1988) 参照。O-RS、O-tRNA 及び突然変異体遺伝子を *in vivo* 融合すると、UAG コドンに 10
応答して非天然アミノ酸が組込まれ、特定位置に非天然アミノ酸を含むタンパク質が得られる。

【0133】

非天然アミノ酸の *in vivo* 組込みは宿主（例えば大腸菌）にさほど影響を与えずに行うことができる。例えば、UAG コドンの抑圧効率は O-tRNA（例えばアンバーサプレッサー tRNA）と（UAG コドンに結合してリボソームから増殖ペプチドの放出を開始する）放出因子 1（RF1）の競合に依存するので、抑圧効率は、例えば O-tRNA（例えばサプレッサー tRNA）の発現レベルを高めるか又は RF1 欠損株を使用することにより調節することができる。

【0134】

非天然アミノ酸はレアコドンでコードすることもできる。例えば、*in vitro* タンパク質合成反応でアルギニン濃度を下げると、レアアルギニンコドン AGG はアラニンでアシル化された合成 tRNA による Ala の挿入に有効であることが分かっている。例えば C. H. Ma, W. Kudlicki, O. W. Odom, G. Kramer and B. Hardesty, *Biochemistry*, 32: 7939 (1993) 参照。この場合には、合成 tRNA は大腸菌に少量種として存在する天然 tRNA Arg と競合する。生物によっては全三重項コドンを使用しないものもある。Micrococcus luteus で割り当てられないコドン AGA が *in vitro* 転写／翻訳抽出物へのアミノ酸の挿入に使用されている。例えば A. K. Kowal と J. S. Oliver, *Nucl. Acid. Res.*, 25: 4685 (1997) 参照。本発明の成分は 30
これらのレアコドンを *in vivo* 使用するために作製することができる。

【0135】

セクターコドンは更に 4 塩基以上のコドン（例えば 4、5、6、又は 7 塩基以上のコドン）も含む。4 塩基コドンの例としては例えば AGGA、CUAG、UAGA、CCCU 等が挙げられる。5 塩基コドンの例としては例えば AGGAC、CCCCU、CCUC、CUAGA、CUACU、UAGGC 等が挙げられる。例えば、突然変異 O-tRNA（例えば特殊フレームシフトサプレッサー tRNA）とアンチコドンループ（例えば少なくとも 8~10 nt アンチコドンループ）の存在下では、4 塩基以上のコドンは単一アミノ酸として読取られる。他の態様では、アンチコドンループは例えば少なくとも 4 塩基コドン、少なくとも 5 塩基コドン、又は少なくとも 6 塩基コドン又はそれ以上をデコード 40
することができる。4 塩基コドンは 256 種が考えられるので、4 塩基以上のコドンを使用すると同一細胞で多重非天然アミノ酸をコードすることができる。J. Christopher Anderson ら, *Exploring the Limits of Codon and Anticodon Size*, *Chemistry and Biology*, Vol. 9, 237-244 (2002); Thomas J. Magliery, *Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in Escherichia coli*, *J. Mol. Biol.* 307: 755-769 (50

2001) 参照。

【0136】

本発明の方法はフレームシフト抑圧に基づく拡張コドンを使用する。4塩基以上のコドンは例えば1個又は多数の非天然アミノ酸を同一タンパク質に挿入することができる。例えば、*in vitro* 生合成法を使用して非天然アミノ酸をタンパク質に組込むために4塩基コドンが使用されている。例えばC. H. Ma, W. Kudlicki, O. W. Odom, G. Kramer and B. Hardesty, *Biochemistry*, 1993, 32, 7939 (1993); 及びT. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami and M. Sisido, *J. Am. Chem. Soc.*, 121: 34 (1999) 参照。2個の化学的にアシル化したフレームシフトサプレッサートRNAを使用して2-ナフチルアラニンとリジンのNB D誘導体を同時にストレプトアビジンに*in vitro* 組込むためにCGGGとAGGUが使用されている。例えばT. Hohsaka, Y. Ashizuka, H. Sasaki, H. Murakami and M. Sisido, *J. Am. Chem. Soc.*, 121: 12194 (1999) 参照。*in vivo* 研究では、MooreらがNCUAアンチコドンをもつtRNA^{Leu}誘導体がUAGNコドンを抑圧できるかどうかを試験し(NはU、A、G又はCであり得る)、四重項UAGAはUCUAアンチコドンをもつtRNA^{Leu}により13から26%の効率でデコードすることができるが、0又は-1フレームでは殆どデコードできないことを見出した。B. Moore, B. C. Persson, C. C. Nelson, R. F. Gesteland and J. F. Atkins, *J. Mol. Biol.*, 298: 195 (2000) 参照。1態様では、本発明ではレアコドン又はナンセンスコドンに基づく拡張コドンを使用し、他の望ましくない部位でのミスセンス読み飛ばしとフレームシフト抑圧を減らすことができる。

10

20

【0137】

翻訳バイパス系を使用して非天然アミノ酸を所望ポリペプチドに組込むこともできる。1翻訳バイパス系では、大きい配列を遺伝子に挿入するが、タンパク質に翻訳しない。この配列はリボソームに配列を飛び越させて挿入の下流の翻訳を再開するための合図として機能する構造を含む。

【0138】

非天然アミノ酸をポリペプチドに組込むための他の上記方法の代用又は併用してトランス翻訳系を使用することができる。この系は大腸菌に存在するtmRNAと呼ぶ分子を利用する。このRNA分子はアラニルトRNAに構造的に近縁であり、アラニルシンテターゼによりアミノアシル化される。tmRNAとtRNAの相違はアンチコドンループが特殊な大きい配列で置換されていることである。この配列は鋳型としてtmRNA内でコードされるオープンリーディングフレームを使用して停止した配列でリボソームに翻訳を再開させることができる。本発明では、直交シンテターゼで優先的にアミノアシル化して非天然アミノ酸を負荷することができる直交tmRNAを作製することができる。この系を使用する遺伝子を転写することにより、リボソームは特定部位で停止し、非天然アミノ酸はこの部位に導入され、その後、直交tmRNA内にコードされた配列を使用して翻訳が再開する。

30

40

【0139】

セクターコドンは場合により非天然塩基対を含む。これらの非天然塩基対は更に既存遺伝アルファベットを拡張する。塩基対が1個増えると、三重項コドンの数は64から125に増す。3番目の塩基対の性質としては安定的且つ選択的な塩基対合、高い忠実度で効率的なポリメラーゼによるDNAへの酵素組込み、及び新生非天然塩基対の合成後の効率的な連続プライマー伸長が挙げられる。方法と組成物に応用可能な非天然塩基対については、例えばHiraora, *An unnatural base pair for incorporating amino acid analogues into protein*, *Nature Biotechnology*, 20: 177-182 (2002) に記載されている。他の関連文献は以下に挙げる。

50

【0140】

in vivo 使用では、非天然ヌクレオシドは膜浸透性であり、リン酸化して対応する三リン酸塩を形成する。更に、増加した遺伝情報は安定しており、細胞酵素により破壊されない。Benner らによる従来の研究ではカノニカルワトソンクリック対とは異なる水素結合パターンを利用しており、そのうちで最も注目に価する例はイソC：イソG対である。例えばC. Switzer, S. E. Moroney and S. A. Benner, J. Am. chem. Soc., 111: 8322 (1989); J. A. Piccirilli, T. Krauch, S. E. Moroney and S. A. Benner, Nature, 1990, 343: 33 (1990); 及びE. T. Kool, Curr. Opin. Chem. Biol., 4: 602 (2000) 参照。これらの塩基は一般に天然塩基とある程度ミスペアであり、酵素により複製することができない。Kool らは水素結合の代わりに塩基間の疎水性充填相互作用により塩基対の形成を誘導できることを示した。E. T. Kool, Curr. Opin. Chem. Biol., 4: 602 (2000); 及びK. M. GuckianとE. T. Kool, Angew. Chem. Int. Ed. Engl., 36, 2825 (1998) 参照。上記全要件を満足する非天然塩基対を開発する試みとして、Schultz, Romersberg らは一連の非天然疎水性塩基を体系的に合成し、試験している。PICS：PICS 自己対が天然塩基対よりも安定であることが判明し、大腸菌DNAポリメラーゼIのKlenowフラグメント(KF)によりDNAに効率的に組込むことができる。例えばD. L. McMin, A. K. Ogawa, Y. Q. Wu, J. Q. Liu, P. G. Schultz and F. E. Romersberg, J. Am. Chem. Soc., 121: 11586 (1999); 及びA. K. Ogawa, Y. Q. Wu, K. L. McMin, J. Q. Liu, P. G. Schultz and F. E. Romersberg, J. Am. Chem. Soc., 122: 3274 (2000) 参照。生物機能に十分な効率と選択性でKFにより3MN：3MN自己対を合成することができる。例えばA. K. Ogawa, Y. Q. Wu, M. Berger, P. G. Schultz and F. E. Romersberg, J. Am. Chem. Soc., 122: 8803 (2000) 参照。しかし、どちらの塩基も後期複製用チェーンターミネーターとして作用するに止まっている。PICS 自己対を複製するために使用できる突然変異体DNAポリメラーゼが最近開発された。更に、7AI 自己対も複製できる。例えばE. J. L. Tae, Y. Q. Wu, G. Xia, P. G. Schultz and F. E. Romersberg, J. Am. Chem. Soc., 123: 7439 (2001) 参照。Cu(II)と結合すると安定な対を形成する新規メタロ塩基対Dipic：Pyも開発されている。E. Meggers, P. L. Holland, W. B. Tolman, F. E. Romersberg and P. G. Schultz, J. Am. Chem. Soc., 122: 10714 (2000)。拡張コドンと非天然コドンは天然コドンに本質的に直交性であるので、本発明の方法はその直交tRNAを作製するためにこの性質を利用することができる。

【0141】

非天然アミノ酸

本明細書において非天然アミノ酸とはセレノシステインと20種の遺伝的にコードされるαアミノ酸（即ちアラニン、アルギニン、アスパラギン、アスパラギン酸、システイン、グルタミン、グルタミン酸、グリシン、ヒスチジン、イソロイシン、ロイシン、リジン、メチオニン、フェニルアラニン、プロリン、セリン、スレオニン、トリプトファン、チロシン、バリン）以外の任意アミノ酸、改変アミノ酸又はアミノ酸類似体を意味する。αアミノ酸の一般構造は式I：

【0142】

【化1】

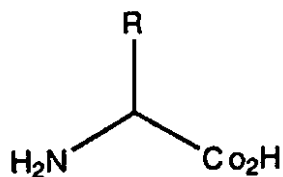
10

20

30

40

I



10

により表される。

【0143】

非天然アミノ酸は一般に式 I をもつ任意構造であり、式中、R 基は 20 種の天然アミノ酸で使用されている以外の任意置換基である。20 種の天然アミノ酸の構造については、例えば L. Stryer 著 Biochemistry, 第 3 版, 1988, Freeman and Company, New York を参照されたい。本発明の非天然アミノ酸は上記 20 種の α アミノ酸以外の天然化合物でもよいことに留意されたい。

【0144】

本発明の非天然アミノ酸は一般に側鎖のみが天然アミノ酸と異なるので、本発明の非天然アミノ酸は天然タンパク質で形成されると同様に他のアミノ酸（例えば天然又は非天然アミノ酸）とアミド結合を形成する。他方、非天然アミノ酸は天然アミノ酸と異なる側鎖基をもつ。例えば、式 I 中の R は場合によりアルキル、アリール、アシル、ケト、アジド、ヒドロキシル、ヒドラジン、シアノ、ハロ、ヒドラジド、アルケニル、アルキニル、エーテル、チオール、セレノ、スルホニル、硼酸、ボロン酸、ホスホ、ホスホノ、ホスフィン、複素環、エノン、イミン、アルデヒド、エステル、チオ酸、ヒドロキシルアミン、アミノ基等又はその任意組合せを含む。他の該当非天然アミノ酸としては、光架橋基を含むアミノ酸、スピンラベル付きアミノ酸、蛍光アミノ酸、金属結合性アミノ酸、金属含有アミノ酸、放射性アミノ酸、新規官能基をもつアミノ酸、他の分子と共有又は非共有的に相互作用するアミノ酸、フォトケージド及び／又は光異性化可能なアミノ酸、ビオチン又はビオチン類似体含有アミノ酸、グリコシル化アミノ酸（例えば糖鎖置換セリン）、他の糖鎖修飾アミノ酸、ケト含有アミノ酸、ポリエチレングリコール又はポリエーテルを含むアミノ酸、重原子置換アミノ酸、化学分解性又は光分解性アミノ酸、天然アミノ酸に比較して延長側鎖（例えばポリエーテル又は例えば炭素長が約 5 もしくは約 10 を越える長鎖炭化水素）をもつアミノ酸、炭素結合糖含有アミノ酸、レドックス活性アミノ酸、アミノチオ酸含有アミノ酸、及び 1 個以上の毒性部分を含むアミノ酸が挙げられるが、これらに限定されない。

20

30

【0145】

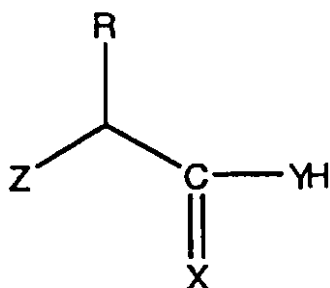
新規側鎖を含む非天然アミノ酸に加え、非天然アミノ酸は場合により例えば式 I I 及び I I I :

40

【0146】

【化 2】

II

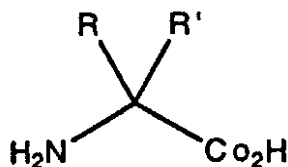


10

【0147】

【化3】

III



20

【0148】

30

の構造により表されるような修飾主鎖構造を含み、式中、Zは一般にOH、NH₂、SH、NH-R'又はS-R'を含み、XとYは同一でも異なってもよく、一般にS又はOであり、RとR'は場合により同一又は異なり、一般に式Iをもつ非天然アミノ酸について上記に記載したR基と同一の基及び水素から選択される。例えば、本発明の非天然アミノ酸は場合により式II及びIIIにより表されるようにアミノ又はカルボキシル基に置換を含む。この種の非天然アミノ酸としては、例えば20種の共通天然アミノ酸に対応する側鎖又は非天然側鎖をもつα-ヒドロキシ酸、α-チオ酸、α-アミノチオカルボキシレートが挙げられるが、これらに限定されない。更に、α-炭素の置換は場合によりL、D又はα, α-ジ置換アミノ酸（例えばD-グルタミン酸、D-アラニン、D-メチル-α-チロシン、アミノ酪酸等）を含む。他の構造としては環状アミノ酸（例えばプロリン類似体や、3、4、6、7、8及び9員環プロリン類似体）、β及びγアミノ酸（例えば置換β-アラニン及びγ-アミノ酪酸）が挙げられる。

40

【0149】

例えば、多数の非天然アミノ酸は天然アミノ酸（例えばチロシン、グルタミン、フェニルアラニン等）に由来する。チロシン類似体としてはパラ置換チロシン、オルト置換チロシン、及びメタ置換チロシンが挙げられ、置換チロシンはアセチル基、ベンゾイル基、アミノ基、ヒドラジン、ヒドロキシルアミン、チオール基、カルボキシ基、イソプロピル基、メチル基、C₆-C₂₀直鎖又は分枝鎖炭化水素、飽和又は不飽和炭化水素、O-メチル基、ポリエーテル基、ニトロ基等を含む。更に、多重置換アリール環も考えられる。本発明のグルタミン類似体としてはα-ヒドロキシ誘導体、γ置換誘導体、環状誘導体及びアミ

50

ド置換グルタミン誘導体が挙げられるが、これらに限定されない。フェニルアラニン類似体の例としてはヒドロキシ基、メトキシ基、メチル基、アリル基、アセチル基等を置換基に含むメタ置換フェニルアラニンが挙げられるが、これらに限定されない。非天然アミノ酸の特定例としてはO-メチル-L-チロシン、L-3-(2-ナフチル)アラニン、3-メチルフェニルアラニン、O-4-アリル-L-チロシン、4-プロピル-L-チロシン、トリ-O-アセチル-GlcNAc β -セリン、L-Dopa、フッ素化フェニルアラニン、イソプロピル-L-フェニルアラニン、p-アジド-L-フェニルアラニン、p-アシル-L-フェニルアラニン、p-ベンゾイル-L-フェニルアラニン、L-ホスホセリン、ホスホノセリン、ホスホノチロシン、p-ヨードフェニルアラニン、p-ブロモフェニルアラニン、p-アミノ-L-フェニルアラニン、及びイソプロピル-L-フェニルアラニン等が挙げられるが、これらに限定されない。各種非天然アミノ酸の構造を図（例えば図17、18、19、26及び29）に示す。

10

【0150】

一般に、本発明の非天然アミノ酸は20種の天然アミノ酸では得られない付加特性を提供するように選択又は設計される。例えば、非天然アミノ酸は場合により（例えばこれらのアミノ酸を組込む）タンパク質の生物学的性質を改変するように設計又は選択される。例えば、非天然アミノ酸をタンパク質に組込むことにより、場合により毒性、生体分布、溶解度、安定性（例えば熱、加水分解、酸化、酵素耐性、分解等）、精製と処理し易さ、構造的性質、分光学的性質、化学及び／又は光化学的性質、触媒活性、酸化還元電位、半減期、他の分子との（例えば共有又は非共有）反応性等の性質を改変する。

20

【0151】

非天然アミノ酸の化学的合成

上記非天然アミノ酸の多くは例えばSigma（米国）やAldrich（Milwaukee, WI, 米国）から市販されている。市販されていないものは場合により下記実施例の記載や当業者に公知の標準方法を使用して合成する。有機合成技術については例えばFessendenとFessenden著Organic Chemistry（1982, 第2版, Willard Grant Press, Boston Mass.）; March著Advanced Organic Chemistry（第3版, 1985, Wiley and Sons, New York）; 及びCareyとSundberg著Advanced Organic Chemistry（第3版, Part A and B, 1990, Plenum Press, New York）を参照されたい。

30

【0152】

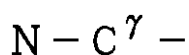
例えば、メタ置換フェニルアラニンは図14に要約するような手順で合成される。一般に、NBS（N-ブロモスクシンイミド）をメタ置換メチルベンゼン化合物に加えてメタ置換臭化ベンジルを得た後にマロン酸化合物と反応させてメタ置換フェニルアラニンを得る。メタ位に使用される典型的置換基としては、ケトン、メトキシ基、アルキル、アセチル等が挙げられるが、これらに限定されない。例えば、3-アセチルフェニルアラニンはNBSを3-メチルアセトフェノンの溶液と反応させることにより製造される。詳細については下記実施例を参照されたい。同様の合成を使用して3-メトキシフェニルアラニンを製造する。その場合の臭化ベンジルのメタ位のR基は-OCH₃である。例えば、Mat

40

【0153】

所定態様では、シンテターゼの活性部位に関する公知情報により非天然アミノ酸の設計を変更し、例えば直交tRNAシンテターゼを使用して直交tRNAをアミノアシル化する。例えば、アミド窒素を置換した誘導体（1）、 γ 位のメチル基（2）、及び

【数1】



環状誘導体 (3) の 3 種のグルタミン類似体が提供される。主要結合部位残基が酵母 GlnRS に相同である大腸菌 GlnRS の X 線結晶構造に基づき、グルタミン側鎖から 10 Å 以内の残基の一連の側鎖突然変異を相補するように類似体を設計し、例えば活性部位 Phe233 から小疎水性アミノ酸への突然変異は Gln の

【数 2】

10



位の立体嵩増加により相補することができる。

【0154】

例えば、N-ナフトロイル-L-グルタミン酸 1, 5-無水物 (図 23 の化合物番号 4) を場合により使用してアミド窒素に置換基をもつグルタミン類似体を合成する。例えば King, F. E. & Kidd, D. A. A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates. J. Chem. Soc., 3315-3319 (1949); Friedman, O. M. & Chatterji, R. Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents. J. Am. Chem. Soc. 81, 3750-3752 (1959); Craig, J. C. ら, Absolute Configuration of the Enantiomers of 7-Chloro-4-[[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine). J. Org. Chem. 53, 1167-1170 (1988); 及び Azoulay, M., Vilmont, M. & Frappier, F. Glutamine analogues as Potential Antimalarials, Eur. J. Med. Chem. 26, 201-5 (1991) 参照。前記無水物は一般にグルタミン酸から製造し、まずアミンをフタルイミドとして保護した後に酢酸中で還流させる。次に無水物を多数のアミンで開環し、アミドに各種置換基を加える。フタロイル基をヒドラジンで脱保護すると、図 23 に示すような遊離アミノ酸が得られる。

20

30

【0155】

γ 位の置換は一般にグルタミン酸のアルキル化により実施される。例えば Koskine, A. M. P. & Rapoport, H. Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues. J. Org. Chem. 54, 1859-1866 (1989) 参照。例えば図 24 の化合物番号 5 に示すような保護アミノ酸は場合によりまずアミノ部分を 9-ブロモ-9-フェニルフルオレン (PhflBr) でアルキル化 (例えば Christie, B. D. & Rapoport, H. Synthesis of Optically Pure Pipecolate from L-Asparagine. Application to the total Synthesis of (+)-Apovincamine throu

40

50

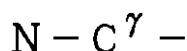
gh Amino Acid Decarbonylation and Iminium Ion Cyclization. J. Org. Chem. 1989, 1859-1866 (1986) 参照) した後、O-tert-ブチル-N, N'-ジイソプロピルイソ尿素を使用して酸部分をエステル化することにより製造される。KN(Si(CH₃)₃)₂を加えてメチルエステルのα位を部位選択的に脱プロトン化してエノラートを形成した後、場合により各種ヨウ化アルキルでアルキル化する。t-ブチルエステルとPhfl基を加水分解すると、所望γ-メチルグルタミン類似体(図24の化合物番号2)が得られた。

【0156】

図25に化合物番号3で示すような

10

【数3】



環状類似体は場合により従来記載されているようにBoc-Asp-Ot-Buから4段階で製造される。例えばBarton, D. H. R., Herve, Y., Potier, P. & Thierry, J. Synthesis of Novel α-Amino- 20
o-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D-α-Amino-Adipic Acids, L-α-aminopimelic Acid and Appropriate Unsaturated Derivatives. Tetradedron Lett. 43, 4297-4308 (1987) 及びSubasinghe, N., Schulte, M., Roon, R. J., Koerner, J. F. & Johnson, R. L. Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropionic acid derivatives and their activity at a novel quisqualate-sensitized site. 30
J. Med. Chem. 35 4602-7 (1992) 参照。N-t-Boc-ピロリジノン、ピロリジノン、又はオキサゾリドンのアニオンの生成後に図25に示すような化合物7を加えるとマイケル付加物が得られる。次にTFAで脱保護すると遊離アミノ酸が得られる。

【0157】

上記非天然アミノ酸以外に、チロシン類似体ライブラリーも設計した。活性部位がM. jannashii シンターゼの活性部位と相高度が高いB. stearoothermophilus TyrRSの結晶構造に基づき、チロシンの芳香族側鎖から10 Å以内の残基を突然変異させた(Y32, G34, L65, Q155, D158, A167, Y32 及びD158)。これらの活性部位アミノ酸の一連の置換を相補するように図26に示 40
すようなチロシン類似体ライブラリーを設計した。これらは種々の疎水性及び水素結合性を付与する各種フェニル置換パターンを含む。チロシン類似体は場合により図27により示す一般ストラテジーを使用して製造される。例えば、場合によりナトリウムエトキシドを使用してアセトアミドマロン酸ジエチルのエノラートを生成する。その後、適当な臭化ベンジルを加えた後に加水分解することにより所望チロシン類似体を製造することができる。

【0158】

非天然アミノ酸の細胞取込み

非天然アミノ酸取込みは例えばタンパク質に組込むために非天然アミノ酸を設計及び選択する場合に一般に検討される問題の1つである。例えば、αアミノ酸は電荷密度が高いの 50

でこれらの化合物が細胞浸透性である可能性は低いと思われる。天然アミノ酸は各種程度のアミノ酸特異性を示す一連のタンパク質輸送系を介して細菌に取込まれる。従って、本発明は非天然アミノ酸が細胞に取込まれるのであるならば、どの非天然アミノ酸が取込まれるかを調べる迅速なスクリーニングを提供する。

【0159】

例えば、場合により各種アミノ酸を最少培地で細胞毒性についてスクリーニングする。毒性は一般に(1)倍加時間の有意変化がない無毒性、(2)倍加時間の増加が約10%未満である低毒性、(3)倍加時間の増加が約10%～約50%である中毒性、(4)倍加時間の増加が約50%～約100%である高毒性、及び(5)倍加時間の増加が約100%を上回る極毒性の5分類に分類される。例えばLiu, D. R. & Schultz, P. G. Progress toward the evolution of an organism with an expanded genetic code. Proceedings of the National Academy of Sciences of the United States of America 96, 4780-4785 (1999) 参照。高毒性又は極毒性に分類されるアミノ酸の毒性は一般にIC50値を得るためにその濃度の関数として測定される。一般に、天然アミノ酸の近縁類似体であるか又は反応性官能性を示すアミノ酸が最高毒性を示す。この傾向はこれらの非天然アミノ酸の毒性メカニズムが天然アミノ酸をもつ必須酵素のタンパク質組込み又は阻害である可能性を示唆している。

10

【0160】

毒性アミノ酸の可能な取込み経路を同定するためには、場合により例えば過剰の構造的に類似する天然アミノ酸を加えた培地でIC50レベルで毒性アッセイを繰返す。毒性アミノ酸では過剰の天然アミノ酸の存在は毒素存在下の細胞の増殖能を救済するが、これは天然アミノ酸が細胞取込み又は必須酵素との結合のために毒素と有効に競合解消するためであると予想される。これらの場合には、場合により毒性アミノ酸を可能な取込み経路に割り当て、細胞生存にその相補が必要な「致死対立遺伝子」と呼ぶ。これらの致死対立遺伝子は細胞が非毒性非天然アミノ酸を取込む能力をアッセイするのに極めて有用である。毒性対立遺伝子の相補は細胞増殖の回復により判断され、恐らく致死対立遺伝子に割り当てられると同一の取込み経路により非毒性アミノ酸が細胞に取込まれることを示唆している。相補の欠如については結論が出ていない。例えば、試験と結論については下記実施例を参照されたい。

20

30

【0161】

(例えば下記実施例に記載するような)得られた結果は、致死非天然アミノ酸対立遺伝子の相補がアミノ酸取込みを定量的に試験するために効率的な方法であることを立証している。この方法は一般に多数の化合物を放射性標識するよりも手間がかからないので所期非天然アミノ酸の分析方法としてより有利な方法である。この一般ストラテジーは場合により核酸塩基類似体、糖鎖類似体又はペプチド類似体等の広範な分子の細胞取込みを迅速に評価するために使用される。例えば、このストラテジーは場合により本明細書に記載する非天然アミノ酸の細胞取込みを評価するために使用される。

【0162】

本発明は、非天然アミノ酸を送達するための一般方法として全アミノ酸取込み経路に独立した方法も提供する。この一般方法は細胞膜を通してジペプチドとトリペプチドを輸送するペプチドパーミアーゼによる取込みに依存する。ペプチドパーミアーゼはあまり側鎖特異的ではなく、その基質に対するKD値はアミノ酸パーミアーゼのKDアミノ酸と同等であり、例えば約0.1mM～約10mMである。例えばNickitenko, A., Trakhanov, S. & Quirocho, S. A structure of DppA, a periplasmic dipeptide transport/chemosensory receptor. Biochemistry 34, 16585-16595 (1995) 及びDuntzen, P., Mowbray, S. L. Crystal structure of the dipeptide binding p

40

50

rotein from *Escherichia coli* involved in active transport and chemotaxis. Protein Science 4, 2327-34 (1995) 参照。その後、非天然アミノ酸はリジン等の天然アミノ酸のコンジュゲートとして取込まれ、内在大腸菌ペプチダーゼの1種によるジペプチドの加水分解後に細胞質に放出される。このアプローチを試験するために、本発明者らは数種のUnn-Lys及びLys-Unnジペプチドを固相合成により合成し、これらのジペプチドの存在下と不在下にリジン最少培地でリジン生合成欠損大腸菌株の増殖を試験した。これらの細胞に利用可能な唯一のリジン源は非天然アミノ酸を含むジペプチドである。ホスホノセリン、ホスホノチロシン、ペンタフルオロフェニルアラニン及びケーゾセリンの取込みをこのように分析した。4例のいずれでも10mM以上のジペプチド濃度で増殖が観察された。取込みは本明細書に記載する方法で容易に分析されるが、細胞取込み経路に利用可能な非天然アミノ酸を設計する別法は、アミノ酸を *in vivo* 生産する生合成経路の提供である。

【0163】

非天然アミノ酸の生合成

細胞にはアミノ酸と他の化合物を生産するために多数の生合成経路が元々存在している。特定非天然アミノ酸の生合成法は自然界（例えば大腸菌中）では存在しないかもしれないが、本発明はこのような方法を提供する。例えば、非天然アミノ酸の生合成経路は場合により新規酵素を付加するか又は既存大腸菌経路を改変することにより大腸菌で作製される。付加新規酵素は場合により天然酵素でも人工的に作製した酵素でもよい。例えば、（下記実施例に記載するような）p-アミノフェニルアラニンの生合成は他の生物からの公知酵素の組合せの付加に依存している。これらの酵素の遺伝子はこの遺伝子を含むプラスミドで細胞（例えば大腸菌細胞）を形質転換することにより細胞に導入することができる。遺伝子は細胞で発現されると、所望化合物を合成するための酵素経路を提供する。場合により付加される酵素種の例は下記実施例に記載する。付加酵素配列は例えばGenbankに登録されている。場合により人工的に作製した酵素も同様に細胞に付加する。このように、非天然アミノ酸を生産するように細胞機構と細胞資源を操作する。

【0164】

生合成経路で使用する新規酵素の生産方法又は既存経路を進化させる方法は種々のものが入手可能である。例えば、場合により例えばMaxygen, Inc. (maxygen.comの世界ウェブサイト)により開発されたような再帰的組換えを使用して新規酵素及び経路を開発する。例えば、Stemmer 1994, "Rapid evolution of a protein in vitro by DNA shuffling," Nature Vol. 370 No. 4: Pg. 389-391; 及びStemmer, 1994, "DNA Shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution," Proc. Natl. Acad. Sci. USA. Vol. 91: Pg. 10747-10751 参照。同様に、場合によりGenencor (genencor.comの世界ウェブサイト)により開発されたDesignPath (登録商標)を代謝経路組換えに使用し、例えば大腸菌でO-メチル-L-チロシンを生産するように経路を組換える。この技術は例えば機能ゲノミクスと分子進化及び設計により同定した新規遺伝子の組合せを使用して宿主生物で既存経路を再構成する。Diversa Corporation (diversa.comの世界ウェブサイト)も例えば新規経路を作製するために遺伝子ライブラリーと遺伝子経路を迅速にスクリーニングするための技術を提供している。

【0165】

一般に、本発明の生合成方法（例えばコリスミン酸からp-アミノフェニルアラニン (pAF)を生産するための経路）は細胞で生産される他のアミノ酸濃度に影響を与えない。例えば、コリスミン酸からpAFを生産するために使用される経路は細胞内でpAFを生産するが、コリスミン細胞から一般に生産される他の芳香族アミノ酸の濃度は実質的に変

わらない。一般に本発明の組換え生合成経路で生産される非天然アミノ酸は効率的タンパク質生合成に十分な濃度（例えば天然細胞量）で生産されるが、他のアミノ酸の濃度を变化させたり細胞資源を枯渇させる程にはならない。こうして *in vivo* 生産される典型濃度は約 10 mM～約 0.05 mM である。特定経路に所望される酵素を生産するために使用される遺伝子を含むプラスミドで細菌を形質転換し、21番目のアミノ酸（例えば pAF、dopa、O-メチル-L-チロシン等）を作製したら、場合により *in vivo* 選択を使用してリボソームタンパク質合成と細胞増殖の両者に合うように非天然アミノ酸の生産を更に最適化させる。

【0166】

非天然アミノ酸を組み込んだタンパク質を含む組成物

10

本発明は少なくとも1種の非天然アミノ酸を組み込んだタンパク質を含む組成物を提供する。本発明は本発明の組成物と方法を使用して生産した少なくとも1種の非天然アミノ酸を組み込んだタンパク質を含む組成物も提供する。1態様では、タンパク質を細胞依存的にプロセッシング及び改変（例えばリン酸化、グルコシル化、折り畳み、膜結合等）する。

【0167】

1側面では、組成物は場合により少なくとも約 10 µg、例えば少なくとも約 50 µg、少なくとも約 100 µg、少なくとも約 500 µg、少なくとも約 1 mg、又は少なくとも約 10 mg 以上、例えば *in vivo* タンパク質生産法で達成可能な量のタンパク質を含む（組換えタンパク質の生産と精製の詳細については本明細書に記載する）。例えば、タンパク質は場合により（例えば約 1 nL～約 100 L の任意容量の）例えば細胞溶解液、医薬緩衝液又は他の液体懸濁液中に少なくとも 10 µg/L、少なくとも約 50 µg/L、少なくとも約 100 µg/L、少なくとも約 500 µg/L、少なくとも約 1 mg/L、又は少なくとも約 10 mg/L 以上の濃度で組成物に存在する。少なくとも1種の非天然アミノ酸を含むタンパク質を大量（例えば他の方法、例えば *in vitro* 翻訳で一般に可能な量よりも多い量）に生産することが本発明の1つの特徴であり、従来技術にまさる効果である。

20

【0168】

少なくとも1種の非天然アミノ酸を含むタンパク質を大量（例えば他の方法、例えば *in vitro* 翻訳で一般に可能な量よりも多い量）に生産することは本発明の1つの特徴であり、従来技術にまさる効果である。例えば、重金属を含むタンパク質を大量に合成できるので、例えばX線結晶解析によるタンパク質構造の決定が容易になる。

30

【0169】

非天然アミノ酸の組み込みは例えば寸法、酸性度、求核性、水素結合、疎水性、プロテアーゼ標的部位接近性等を改変するように例えばタンパク質構造及び／又は機能をカスタマイズするために実施することができる。非天然アミノ酸を含むタンパク質は触媒性又は物性を強化するか又は全く新規にすることができる。例えば、非天然アミノ酸をタンパク質に組み込むことにより、場合により毒性、生体分布、構造的性質、分光学的性質、化学及び／又は光化学的性質、触媒能、半減期（例えば血清半減期）、他の分子との（例えば共有又は非共有）反応性等の性質を改変する。少なくとも1種の非天然アミノ酸を組み込んだタンパク質を含む組成物は例えば新規治療、診断、触媒酵素、結合タンパク質（例えば抗体）、及び例えばタンパク質構造と機能の研究に有用である。

40

【0170】

本発明の1側面では、組成物は少なくとも1種、例えば少なくとも2種、少なくとも3種、少なくとも4種、少なくとも5種、少なくとも6種、少なくとも7種、少なくとも8種、少なくとも9種、少なくとも10種、又は11種以上の非天然アミノ酸を組み込んだ少なくとも1種のタンパク質を含む。2種以上の非天然アミノ酸を組み込んだ所与タンパク質では、非天然アミノ酸は同一でも異なってもよい（例えばタンパク質は2種以上の異なる型の非天然アミノ酸を組み込んでよいし、非天然アミノ酸をもつ2個以上の異なる部位を組み込んでよいし、その両方でもよい）。

【0171】

50

非天然アミノ酸を組込んだほぼ任意タンパク質（及び例えば1種以上のセレクターコドン
を組込んだ対応する任意コーディング核酸）は本明細書に記載する組成物と方法を使用し
て生産することができる。数十万種の公知タンパク質を同定しようとするのではなく、例
えば該当翻訳系に1種以上の適当なセレクターコドンを組込むように入手可能な任意突然
変異法をカスタマイズすることにより、1種以上の非天然アミノ酸を組込むように公知タ
ンパク質の任意のものを改変することができる。公知タンパク質の一般的な配列寄託機関
としてはGenBank、EMBL、DDBJ及びNCBIが挙げられる。他の寄託機関
もインターネットを検索することにより容易に確認できる。

【0172】

本明細書に記載する非天然アミノ酸を *in vivo* 組込むための組成物と方法を使用し 10
て生産することができるタンパク質の1好適類としては治療用タンパク質が挙げられる。
1種以上の非天然アミノ酸を組込むように改変可能な治療用及び他のタンパク質の例とし
ては例えば α 1アンチトリプシン、アンジオスタチン、抗血友病因子、抗体（抗体の詳細
については後述する）、アポリポタンパク質、アポタンパク質、心房性ナトリウム利尿因
子、心房性ナトリウム利尿ポリペプチド、心房性ペプチド、C-X-Cケモカイン（例え
ばT39765、NAP-2、ENA-78、Gro-a、Gro-b、Gro-c、I
P-10、GCP-2、NAP-4、SDF-1、PF-4、MIG）、カルシトニン、
CCケモカイン（例えば単球化学誘引タンパク質-1、単球化学誘引タンパク質-2、単
球化学誘引タンパク質-3、単球炎症性タンパク質-1 α 、単球炎症性タンパク質-1 β 20
、RANTES、I309、R83915、R91733、HCC1、T58847、D
31065、T64262）、CD40リガンド、Cキットリガンド、コラーゲン、コロ
ニー刺激因子（CSF）、補体因子5a、補体阻害剤、補体受容体1、サイトカイン（例
えば上皮好中球活性化ペプチド-78、GRO α /MGSA、GRO β 、GRO γ 、MI
P-1 α 、MIP-1 δ 、MCP-1）、表皮増殖因子（EGF）、エリスロポエチン（
「EPO」、1種以上の非天然アミノ酸の組込みによる改変の好適ターゲット）、表皮剥
離毒素A及びB、IX因子、VII因子、VII因子、X因子、繊維芽細胞増殖因子（
FGF）、フィブリノーゲン、フィブロンectin、G-CSF、GM-CSF、グルコセ
レブロシダーゼ、ゴナドトロピン、増殖因子、ヘッジホッグタンパク質（例えばソニック
、インディアン、デザート）、ヘモグロビン、肝細胞増殖因子（HGF）、ヒルジン、ヒ
ト血清アルブミン、インスリン、インスリン様増殖因子（IGF）、インターフェロン（ 30
例えばIFN- α 、IFN- β 、IFN- γ ）、インターロイキン（例えばIL-1、I
L-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、I
L-10、IL-11、IL-12等）、ケラチノサイト増殖因子（KGF）、ラクトフ
ェリン、白血病阻害因子、ルシフェラーゼ、ニールチュリン、好中球阻害因子（NIF
）、オンコスタチンM、骨形成タンパク質、副甲状腺ホルモン、PD-ECSF、PDG
F、ペプチドホルモン（例えばヒト増殖ホルモン）、プレイオトロピン、プロテインA、
プロテインG、発熱外毒素A、B及びC、リラキシン、レニン、SCF、可溶性補体受容
体I、可溶性I-CAM1、可溶性インターロイキン受容体（IL-1、2、3、4、5
、6、7、8、9、10、11、12、13、14、15）、可溶性TNF受容体、ソマ
トメジン、ソマトスタチン、ソマトトロピン、ストレプトキナーゼ、スーパー抗原即ちブ 40
ドウ球菌エンテロトキシン（SEA、SEB、SEC1、SEC2、SEC3、SED、
SEE）、スーパーオキシドジスムターゼ、毒素性ショック症候群毒素（TSST-1）
、チモシン α 1、組織プラスミノゲンアクチベーター、腫瘍壊死因子 β （TNF β ）、
腫瘍壊死因子受容体（TNFR）、腫瘍壊死因子 α （TNF α ）、血管内皮増殖因子（V
EGEF）、ウロキナーゼ等が挙げられる。

【0173】

これらのタンパク質の多くは市販されており（例えばSigma BioScience
s 2002カタログ及び価格表参照）、対応するタンパク質配列と遺伝子及び、一般に
多くのその変異体は周知である（例えばGenebank参照）。例えば1種以上の所期
治療特性に関してタンパク質を改変するように本発明に従って1種以上の非天然アミノ酸 50

を挿入することにより前記タンパク質の任意のものを改変することができる。治療関連特性の例としては血清半減期、貯蔵半減期、安定性、免疫原性、治療活性、（例えば非天然アミノ酸へのレポーター基（例えばラベル又はラベル結合部位）の組込みによる）検出性、LD-50又は他の副作用の低減、胃を通して体内に導入できること（例えば経口利用性）等が挙げられる。

【0174】

本明細書に記載する非天然アミノ酸を *in vivo* 組込むための組成物と方法を使用して生産することができるタンパク質の1類としては転写及び発現アクチベーターが挙げられる。転写及び発現アクチベーターの例としては細胞増殖、分化、制御等を調節する遺伝子とタンパク質が挙げられる。発現及び転写アクチベーターは原核生物、ウイルス及び真核生物（例えば真菌類、植物及び哺乳動物を含む動物）に存在し、広範な治療ターゲットとなる。自明の通り、発現及び転写アクチベーターは例えば受容体との結合、シグナル導入カスケードの刺激、転写因子の発現調節、プロモーターやエンハンサーとの結合、プロモーターやエンハンサーと結合するタンパク質との結合、DNA巻き戻し、pre-mRNA スプライシング、RNA ポリアデニル化及びRNA 分解等の多数のメカニズムにより転写を調節する。

10

【0175】

本発明のタンパク質（例えば1種以上の非天然アミノ酸を組込んだタンパク質）の1好適類としては発現アクチベーター（例えばサイトカイン、炎症性分子、増殖因子、その受容体、及び腫瘍遺伝子産物、例えばインターロイキン（例えばIL-1、IL-2、IL-8等）、インターフェロン、FGF、IGF-I、IGF-II、FGF、PDGF、TNF、TGF- α 、TGF- β 、EGF、KGF、SCF/c-キット、CD40L/CD40、VLA-4/VCAM-1、ICAM-1/LFA-1及びヒアルリン/CD44）と、シグナル伝達分子及び対応する腫瘍遺伝子産物（例えばMos、Ras、Raf及びMet）と、転写アクチベーター及びサプレッサー（例えばp53、Tat、Fos、Myc、Jun、Myb、Rel、及びステロイドホルモン受容体（例えばエストロゲン、プロゲステロン、テストステロン、アルドステロン、LDL受容体リガンド及びコレチコステロン））が挙げられる。

20

【0176】

他の各種タンパク質も本発明の1種以上の非天然アミノ酸を組込むように改変することができる。例えば、本発明は1種以上のワクチンタンパク質における1種以上の天然アミノ酸を非天然アミノ酸で置換することができ、このようなタンパク質は例えば感染性真菌類（例えばAspergillus、Candida種）、細菌類、特に病原細菌モデルとして利用できる大腸菌や医学的に重要な細菌（例えばStaphylococci（例えばaureus）又はStreptococci（例えばpneumoniae））、原生動物（例えば胞子虫類（例えばPlasmodia）、根足虫類（例えばEntamoeba）及び鞭毛虫類（Trypanosoma、Leishmania、Trichomonas、Giardia等））、ウイルス（例えば（+）RNAウイルス（例えばポックスウイルス（例えばワクシニア）、ピコルナウイルス（例えばポリオ）、トガウイルス（例えば風疹）、フラビウイルス（例えばHCV）、及びコロナウイルス）、（-）RNAウイルス（例えばラブドウイルス（例えばVSV）、パラミクソウイルス（例えばRSV）、オルトミクソウイルス（例えばインフルエンザ）、プンヤウイルス及びアレナウイルス）、dsDNAウイルス（例えばレオウイルス）、RNA→DNAウイルス（即ちレトロウイルス、例えばHIV及びHTLV）、及び所定のDNA→RNAウイルス（例えばB型肝炎））に由来する。

30

40

【0177】

本明細書に記載する方法に従って1種以上の非天然アミノ酸を組込むように各種酵素（例えば産業用酵素）を改変することもでき、このような酵素としてはアミダーゼ、アミノ酸ラセマーゼ、アシラーゼ、デハロゲナーゼ、ジオキシゲナーゼ、ジアリールプロパンペルオキシダーゼ、エピメラーゼ、エポキシドヒドロラーゼ、エステラーゼ、イソメラーゼ、

50

キナーゼ、グルコースイソメラーゼ、グリコシダーゼ、グリコシルトランスフェラーゼ、ハロペルオキシダーゼ、モノオキシゲナーゼ（例えば p 450 類）、リパーゼ、リグニンペルオキシダーゼ、ニトリルヒドラーターゼ、ニトリラーゼ、プロテアーゼ、ホスファターゼ、スブチリシン、トランスアミナーゼ及びヌクレアーゼが挙げられる。

【0178】

昆虫耐性タンパク質（例えば C r y タンパク質）、澱粉及び脂質生産酵素、植物及び昆虫毒素、毒素耐性タンパク質、マイコトキシン解毒タンパク質、植物成長酵素（例えばリブロース 1, 5-ビスホスフェートカルボキシラーゼ／オキシゲナーゼ、「R U B I S C O」）、リボキシゲナーゼ（L O X）及びホスホエノールピルベート（P E P）カルボキシラーゼ等の農業関連タンパク質も非天然アミノ酸改変に利用可能なターゲットである。

10

【0179】

少なくとも 1 種の非天然アミノ酸を組み込んだタンパク質をコードする遺伝子は本明細書の「一般分子生物学技術」の項に記載する当業者に周知の方法を使用して突然変異誘発することができる。例えば、1 種以上のセクターコドンを組み込むように所期タンパク質の核酸を突然変異誘発し、1 種以上の非天然アミノ酸を挿入できるようにする。本発明は例えば少なくとも 1 種の非天然アミノ酸を組み込んだ任意タンパク質のこのような任意変異体（例えば突然変異体、変形）を含む。

【0180】

同様に、本発明は対応する核酸、即ち 1 種以上の非天然アミノ酸をコードする 1 種以上のセクターコドンを組み込んだ任意核酸も含む。

20

【0181】

1 態様において、本発明は本発明の組成物と方法により生産されたクロラムフェニコールアセチルトランスフェラーゼ（C A T）の A s p 112 T A G 突然変異体を含む組成物を提供し、C A T タンパク質は少なくとも 1 種の非天然アミノ酸（例えば O-メチル-L-チロシン、L-3-（2-ナフチル）アラニン、アミノ、イソプロピル又はアシル含有チロシン類似体等）を含み、タンパク質は組成物中に少なくとも約 100 μ g / L の濃度で存在する。別の態様では、本発明はマウスジヒドロ葉酸レダクターゼ（D H F R）の T y r 163 T A G 突然変異体を含み、D H F R タンパク質は少なくとも 1 種の非天然アミノ酸（例えば O-メチル-L-チロシン、L-3-（2-ナフチル）アラニン、アミノ、イソプロピル又はアシル含有チロシン類似体等）を含み、タンパク質は組成物中に少なくとも約 100 μ g / L の濃度で存在する。

30

【0182】

非天然アミノ酸を含むタンパク質に対する抗体の製造

1 側面では、本発明は非天然アミノ酸及び非天然アミノ酸を含むタンパク質に対する抗体を提供する。非天然アミノ酸及びこのような非天然アミノ酸を含むタンパク質に対する抗体は例えば本発明のタンパク質と非天然アミノ酸を精製するために、精製試薬として有用である。更に、抗体は例えば非天然アミノ酸又は非天然アミノ酸を含むタンパク質の（例えば i n v i v o 又は i n s i t u）存在又は位置を追跡するために、非天然アミノ酸又は非天然アミノ酸を含むタンパク質の存在を指示するための指示試薬として使用することができる。当然のことながら、非天然アミノ酸自体に 1 種以上の非天然アミノ酸を組み込み、1 種以上の非天然アミノ酸により付与される 1 種以上の特性をもつ抗体を提供することもできる。

40

【0183】

本発明の抗体は免疫グロブリン遺伝子又は免疫グロブリン遺伝子のフラグメントにより実質的又は部分的にコードされる 1 種以上のポリペプチドを含むタンパク質とすることができる。認識される免疫グロブリン遺伝子としては κ 、 λ 、 α 、 γ 、 δ 、 ϵ 及び μ 定常領域遺伝子と、無数の免疫グロブリン可変領域遺伝子が挙げられる。L 鎖は κ 又は λ に分類される。H 鎖は γ 、 μ 、 α 、 δ 又は ϵ に分類され、夫々免疫グロブリン I g G、I g M、I g A、I g D 及び I g E 類と言う。典型的な免疫グロブリン（例えば抗体）構造単位は四量体からなる。各四量体は各々 1 個の「L」鎖（約 25 k D）と 1 個の「H」鎖（約 50

50

～70 kD) からなる同一の2対のポリペプチド鎖から構成される。各鎖のN末端は主に抗原認識に関与する約100～110又はそれ以上のアミノ酸からなる可変領域を規定する。可変L鎖(VL)及び可変H鎖(V_H)なる用語は夫々これらのL鎖とH鎖を意味する。

【0184】

抗体は無傷の免疫グロブリンとして存在するか又は各種ペプチドによる消化により生産された多数の明確に特性決定されたフラグメントとして存在する。従って、例えば、ペプシンはヒンジ部のジスルフィド結合下の抗体を消化し、それ自体ジスルフィド結合によりV_H-C_H1に結合したL鎖であるFabの二量体F(ab')₂を生産する。F(ab')₂を温和な条件下に還元してヒンジ部のジスルフィド結合を切断すると、F(ab')₂二量体をFab'単量体に変換することができる。Fab'単量体は主にヒンジ部を構成するFabである(他の抗体フラグメントの更に詳細な説明についてはFundamental Immunology, W. E. Paul編, Raven Press, N. Y. (1999) 参照)。各種抗体フラグメントが無傷の抗体の消化について定義されているが、当業者に自明の通り、化学的又は組換えDNA技術を使用することによりこのようなFab'フラグメント等をde novo合成してもよい。従って、本明細書において抗体なる用語は場合により全抗体の改変により生産されるか又は組換えDNA技術を使用してde novo合成される抗体フラグメントを含む。抗体は単鎖抗体を含み、可変H鎖と可変L鎖が(直接又はペプチドリンカーを介して)結合して連続ポリペプチドを形成している側鎖Fv(sFv又はscFv)抗体を含む。本発明の抗体は例えばポリクローナル、モノクローナル、キメラ、ヒト化、単鎖、Fabフラグメント、Fab発現ライブラリーにより生産されるフラグメント等とすることができる。

【0185】

一般に、本発明の抗体は各種分子生物学又は医薬プロセスで一般試薬及び治療試薬として有用である。ポリクローナル抗体とモノクローナル抗体の製造方法は入手可能であり、本発明の抗体の製造に適用することができる。多数の基礎教科書が標準抗体製造法を記載しており、例えばBorrebaeck(編)(1995)Antibody Engineering, 第2版, Freeman and Company, NY(Borrebaeck); McCaffertyら(1996)Antibody Engineering. A practical approach IRL, Oxford Press, Oxford, England(McCafferty); Paul(1995)Antibody Engineering Protocols Humana Press, Towata, NJ(Paul); Paul(編), (1999)Fundamental Immunology, 第5版, Raven Press, N. Y.; Coligan(1991)Current Protocols in Immunology Wiley/Greene, NY; HarlowとLane(1989)Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY; Stitesら(編)Basic and Clinical Immunology(第4版) Lange Medical Publications, Los Altos, CAとその引用文献; Goding(1986)Monoclonal Antibodies: Principles and Practice(第2版) Academic Press, New York; 及びKohlerとMilstein(1975)Nature 256: 495-497が挙げられる。

【0186】

例えば動物への抗原注射に依存しない組換えによる各種抗体製造技術が開発されており、本発明の関連で使用することができる。例えば、ファージ又は同様のベクターで組換え抗体のライブラリーを作製及び選択することが可能である。例えばWinterら(1994)"Making Antibodies by Phage Display Technology" Annu. Rev. Immunol. 12: 433-55とその引用文献参照。更にGriffithsとDuncan(1998)"Strategies

for selection of antibodies by phage display" Curr Opin Biotechnol 9:102-8; Hoogenboomら (1998) "Antibody phage display technology and its applications" Immunotechnology 4:1-20; Gramら (1992) "in vivo selection and affinity maturation of antibodies from a native combinatorial immunoglobulin library" PNAS 89:3576-3580; Huseら (1989) Science 246:1275-1281; 及び Wardら (1989) Nature 341:544-546 も参照されたい。

10

【0187】

1 態様では、抗体ライブラリーは糸状バクテリオファージの表面にH鎖及びL鎖可変領域を組合せて提示するためにクローニングされる（例えばリンパ球集団から回収するか又は in vitro 組立てた）V遺伝子のレパートリーを含むことができる。ファージは抗原との結合により選択する。ファージ感染細菌から可溶性抗体を発現させ、例えば突然変異誘発により抗体を改良することができる。例えば Balint と Larrick (1993) "Antibody Engineering by Parsimonious Mutagenesis" Gene 137:109-118; Stemmerら (1993) "Selection of an Active Single Chain Fv Antibody From a Protein Linker Library Prepared by Enzymatic Inverse PCR" Bio 20 techniques 14(2):256-65; Cramer i ら (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" Nature Medicine 2:100-103; 及び Cramer i と Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" BioTechniques 18:194-195 参照。

【0188】

組換え抗体ファージ系のクローニング及び発現用キットも公知で市販されており、例えば Amer sham-Pharmacia Biotechnology (Uppsala, スウェーデン) から「組換えファージ抗体系、マウス ScFv モジュール」が販売されている。鎖シャフリングにより高親和性ヒト抗体を製造するためのバクテリオファージ抗体ライブラリーも作製されている（例えば Marks ら (1992) "By-Passing Immunization: Building High Affinity Human Antibodies by Chain Shuffling" Bio techniques 10:779-782 参照）。実際に、抗体は一般に Peptidogenic (pkim@ccnet.com)、HTI Bio-products, inc. (www.htibio.com)、BMA Biomedicals Ltd (40 英国)、Bio-Synthesis, Inc., Research Genetics (Huntsville, Alabama) 等の各種製造業者からオーダーメイドすることができる。

【0189】

所定態様では、例えば抗体を治療投与しようとする場合には、本発明の抗体を「ヒト化」することが有用である。ヒト化抗体を使用すると、（例えば患者がヒトである場合に）治療用抗体に対する望ましくない免疫応答が発生しにくくなる。上記抗体関連文献はヒト化ストラテジーを記載している。ヒト化抗体に加え、ヒト抗体も本発明の特徴である。ヒト抗体はヒト免疫グロブリン配列から構成されることを特徴とする。ヒト抗体は多様な方法を使用して製造することができる（例えば Larrick ら, 米国特許第 5, 001, 0 50

65号参照)。トリオーマ法によりヒト抗体を製造するための一般アプローチはOstbergら(1983), Hybridoma 2:361-367、Ostberg, 米国特許第4,634,664号、及びEngelmanら, 米国特許第4,634,666号に記載されている。

【0190】

タンパク質の精製と検出に抗体を使用する方法は各種のものが公知であり、本明細書に記載するような非天然アミノ酸を含むタンパク質の検出と精製に適用することができる。一般に、抗体はELISA、ウェスタンブロッティング、免疫化学、アフィニティークロマトグラフィー法、SPR及び他の多数の方法の有用な試薬である。上記文献にはELISAアッセイ、ウェスタンブロッティング、表面プラズモン共鳴(SPR)等の実施方法が詳細に記載されている。 10

【0191】

本発明の1側面では、本発明の抗体はそれ自体非天然アミノ酸を含み、所期特性(例えば半減期、安定性、毒性等の改善)をもつ抗体を提供する。抗体は臨床試験で現在使用されている全化合物のほぼ50%を占めており(Witttrup, (1999) "Phage on display" Tibtech 17:423-424)、抗体は診断試薬として広く使用されている。従って、抗体を非天然アミノ酸で改変できるならば、これらの有用試薬を改変するための重要なツールとなる。

【0192】

例えば、MAbは診断分野に多数の用途がある。アッセイは単純なスポット試験から腫瘍イメージングに使用されているDuPont Merck Co. 製品放射標識NR-LU-10MAb(Ruschら(1993) "NR-LU-10 monoclonal antibody scanning. A helpful new adjunct to computed tomography in evaluating non-small-cell lung cancer." J Thorac Cardiovasc Surg 106:200-4)等のより複雑な方法まで多岐にわたる。周知の通り、MAbはELISA、ウェスタンブロッティング、免疫化学、アフィニティークロマトグラフィー法等の中心試薬である。1種以上の非天然アミノ酸を組込むようにこのような任意診断用抗体を改変し、例えばターゲットに対するAbの特異性又はアビディティを改変したり、例えば検出可能なラベル(例えば分光、蛍光、発光等)を非天然アミノ酸に組込むことにより1種以上の検出可能な特性を改変することができる。 20 30

【0193】

有用な抗体試薬の1類は治療用Abである。例えば、抗体は抗体依存性細胞傷害性(ADCC)又は補体介在性溶解(CML)により破壊するために腫瘍細胞にターゲティングすることにより腫瘍増殖を阻止する腫瘍特異的MAbとすることができる(これらの一般型のAbを「魔法の弾丸」とも言う)。1例は非ホジキンリンパ腫の治療に使用される抗CD20MAbであるリツキサンである(Scott(1998) "Rituximab: a new therapeutic monoclonal antibody for non-Hodgkin's lymphoma" Cancer Pract 6:195-7)。第2の例は腫瘍増殖の重要成分を妨害する抗体に関連する。ハーセプチンは転移性乳癌の治療に使用される抗HER-2モノクローナル抗体であり、この作用メカニズムをもつ抗体の1例である(Baselgaら(1998) "Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and doxorubicin against HER2/neu overexpressing human breast cancer xenografts" Cancer Res (1999) 59(8):2020に誤載], Cancer Res 58:2825-31)。 40

【0194】

第3の例は腫瘍又は他の所期部位に直接細胞毒性化合物(毒素、放射性核種等)を送達す 50

るための抗体に関連する。例えば、1例のMabは前立腺腫瘍細胞に放射線を直接送達する90Y結合抗体であるCYT-356である(Debra(1996)"Treatment of hormone-refractory prostate cancer with 90Y-CYT-356 monoclonal antibody"Clin Cancer Res 2:1289-97)。第4の例は抗体特異的酵素プロドラッグ療法であり、腫瘍に同時局在させた酵素が全身投与したプロドラッグを腫瘍近傍で活性化する。例えば、カルボキシペプチダーゼAに結合した抗ep-CAM1抗体が大腸癌治療用に開発中である(Wolfeら(1999)"Antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1:in vitro and in vivo studies with prodrugs of methotrexate and the thymidylate synthase inhibitors GW1031 and GW1843"Bioconj Chem 10:38-48)。この他に、治療効果のために正常細胞機能を特に阻害するように設計したAb(例えば拮抗薬)もある。1例は急性臓器移植拒絶を緩和するためにJohnson and Johnsonから販売されている抗CD3MAbであるオルソクロンOKT3である(Strateら(1990)"Orthoclone OKT3 as first-line therapy in acute renal allograft rejection"Transplant Proc 22:219-20)。別の類の抗体製品は作動薬である。これらのMabは治療効果のために正常細胞機能を特に強化するように設計されている。例えば、アセチルコリン受容体のMab作動薬が神経治療用に開発中である(Xieら(1997)"Direct demonstration of MuSK involvement in acetylcholine receptor clustering through identification of agonist ScFv"Nat. Biotechnol. 15:768-71)。1種以上の治療特性(特異性、アビディティ、血清半減期等)を強化するために1種以上の非天然アミノ酸を組込むようにこれらの抗体に任意のものを改変することができる。

【0195】

別の類の抗体製品は新規機能を提供する。この類の主な抗体は酵素の触媒能に似せて組換えたIg配列等の触媒抗体である(WentworthとJanda(1998)"Catalytic antibodies"Curr Opin Chem Biol 2:138-44)。例えば、興味深い1適用例では触媒抗体mAb-15A10を使用し、中毒治療のためにコカインをin vivo加水分解している(Metsら(1998)"A catalytic antibody against cocaine prevents cocaine's reinforcing and toxic effects in rats"Proc Natl Acad Sci U S A 95:10176-81)。触媒抗体も1種以上の所期性質を改善するために1種以上の非天然アミノ酸を組込むように改変することができる。

【0196】

非天然アミノ酸を含む組換えタンパク質の精製

本発明のタンパク質(例えば非天然アミノ酸を含むタンパク質、非天然アミノ酸を含むタンパク質に対する抗体等)は当業者に使用されている公知標準手順に従って部分的又は実質的に均質まで精製することができる。従って、本発明のポリペプチドは当分野で周知の多数の方法の任意のものにより回収及び精製することができ、このような方法としては例えば硫酸又はエタノール沈殿、酸又は塩基抽出、カラムクロマトグラフィー、アフィニティーカラムクロマトグラフィー、アニオン又はカチオン交換クロマトグラフィー、ホスホセルロースクロマトグラフィー、疎水性相互作用クロマトグラフィー、ヒドロキシアパタイトクロマトグラフィー、レクチンクロマトグラフィー、ゲル電気泳動等が挙げられる。所望によりタンパク質リフォールディング段階を使用して正しく折畳まれた成熟タンパク

質を製造することができる。高純度が所望される場合には高性能液体クロマトグラフィー（HPLC）、アフィニティークロマトグラフィー又は他の適当な方法を最終精製段階でを使用することができる。1態様では、非天然アミノ酸（又は非天然アミノ酸を含むタンパク質）に対して製造した抗体を精製試薬として使用し、例えば1種以上の非天然アミノ酸を含むタンパク質をアフィニティー精製する。所望に応じて部分的又は均質まで精製したら、ポリペプチドを場合により例えばアッセイ成分、治療試薬又は抗体生産用免疫原として使用する。

【0197】

本明細書に引用する他の文献以外に各種精製／タンパク質フォールディング方法が当分野で周知であり、例えばR. Scopes, Protein Purification, Springer-Verlag, N. Y. (1982); Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press, Inc. N. Y. (1990); Sandana (1997) Bioseparation of Proteins, Academic Press, Inc.; Bollagら (1996) Protein Methods, 第2版, Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook Humana Press, NJ; HarrisとAngal (1990) Protein Purification Applications: A Practical Approach IRL Press at Oxford, Oxford, 英国; HarrisとAngal Protein Purification Methods: A Practical Approach IRL Press at Oxford, Oxford, 英国; Scopes (1993) Protein Purification: Principles and Practice 第3版, Springer Verlag, NY; JansonとRyden (1998) Protein Purification: Principles, High Resolution Methods and Applications, 第2版, Wiley-VCH, NY; 及びWalker (1998) Protein Protocols on CD-ROM Humana Press, NJ; 及びその引用文献に記載されている方法が挙げられる。

【0198】

当然のことながら、当業者に自明の通り、合成、発現及び／又は精製後にタンパク質は該当ポリペプチドの所望コンフォーメーションとは異なるコンフォーメーションをもつことができる。例えば、原核系により生産されるポリペプチドは適正なフォールディングを達成するためにカオトロピック剤に暴露して最適化することが多い。例えば大腸菌に由来する溶解液から精製中に発現されたタンパク質を場合により変性させた後、再生させる。これは例えばグアニジンHCl等のカオトロピック剤でタンパク質を可溶化することにより行われる。

【0199】

一般に、発現されたポリペプチドを変性及び還元した後にポリペプチドを好適コンフォーメーションにリフォールディングすることが望ましいことがある。例えば、グアニジン、尿素、DTT、DTE及び／又はシャペロニンを所期翻訳産物に加えることができる。タンパク質の還元、変性及び再生方法も当業者に周知である（上記文献や、Debinskiら (1993) J. Biol. Chem., 268: 14065-14070; KreitmanとPastan (1993) Bioconj. Chem., 4: 581-585; 及びBuchnerら, (1992) Anal. Biochem., 205: 263-270参照）。例えば、Debinskiらはグアニジン-DTEでの封入体タンパク質の変性と還元について記載している。タンパク質は例えば酸化グルタチオンとL-アルギニンを加えたレドックス緩衝液中でリフォールディングすることができる。リフォールディング試薬を流すか又は他の方法で移動させて1種以上のポリペプチド又は他の発現産物と接触させるか、又はポリペプチドを移動させて試薬と接触させることができる

。

【0200】

核酸及びポリペプチド配列変異体

上記及び下記に記載するように、本発明は核酸ポリヌクレオチド配列及びポリペプチドアミノ酸配列（例えばO-tRNA及びO-RS）と、例えば前記配列を含む組成物と方法を提供する。前記配列（例えばO-tRNA及びO-RS）の例を本明細書に開示する。しかし、当業者に自明の通り、本発明は本明細書（例えば実施例）に開示する配列に限定されない。当業者に自明の通り、本発明は本明細書に記載する機能をもつ（例えば（例えばO-tRNA又はO-RSをコードする））多数の非関連配列も提供する。

【0201】

同様に当業者に自明の通り、開示配列の多数の変異体も本発明に含まれる。例えば、機能的に同一配列である開示配列の保存変異も本発明に含まれる。少なくとも1種の開示配列にハイブリダイスする核酸ポリヌクレオチド配列の変異体も本発明に含むものとする。例えば標準配列比較法により本明細書に開示する配列のユニークサブ配列であるとみなされる配列も本発明に含まれる。

10

【0202】

保存変異

遺伝コードの縮重により、「サイレント置換」（即ちコードされるポリペプチドに変化を生じない核酸配列の置換）はアミノ酸をコードする全核酸配列の暗黙の特徴である。同様に、「保存アミノ酸置換」はアミノ酸配列中の1又は数個のアミノ酸を高度に類似する特性をもつ別のアミノ酸で置換するものであり、このような置換も開示構築物と高度に類似するとは自明である。各開示配列のこのような保存変異は本発明の1つの特徴である。

20

【0203】

特定核酸配列の「保存変異」とは同一又は本質的に同一のアミノ酸配列をコードする核酸を意味し、核酸がアミノ酸配列をコードしない場合には本質的に同一の配列を意味する。当業者に自明の通り、コードされる配列中の単一アミノ酸又は低百分率（一般に5%未満、より一般には4%、2%又は1%未満）のアミノ酸を置換、付加又は欠失させる個々の置換、欠失又は付加の結果としてアミノ酸1個を欠失するか、アミノ酸1個が付加するか、又はアミノ酸1個が化学的に類似するアミノ酸1個で置換される場合には、これらの改変は「保存改変変異」である。従って、本発明に示すポリペプチド配列の「保存変異」はポリペプチド配列のアミノ酸の低百分率、一般に5%未満、より一般には2%又は1%未満を同一保存置換基の保存可能に選択したアミノ酸で置換する置換を含む。最後に、非機能的配列の付加のように核酸分子のコードされる活性を変えない配列の付加も基本核酸の保存変異である。

30

【0204】

機能的に類似するアミノ酸を生じる保存置換表は当分野で周知である。相互に「保存置換」を含む天然アミノ酸を含む基の例を以下に示す。

【0205】

保存置換基

1	アラニン(A) セリン(s) スレオニン(T)
2	アスパラギン酸(D) グルタミン酸(E)
3	アスパラギン(N) グルタミン(Q)
4	アルギニン(R) リジン(K)
5	イソロイシン(I) ロイシン(L) メチオニン(M) バリン(V)
6	フェニルアラニン(F) チロシン(Y) トリプトファン(W)

10

【0206】

20

核酸ハイブリダイゼーション

比較ハイブリダイゼーションを使用して（本発明の核酸の保存変異を含む）本発明の核酸を同定することができ、この比較ハイブリダイゼーション法は本発明の核酸を識別する好適な方法である。更に、高、超高及び超々高ストリンジェンシー条件下で配列番号1～34に示す核酸とハイブリダイズするターゲット核酸も本発明の1つの特徴である。このような核酸の例としては所定核酸配列と比較して1又は数個のサイレント又は保存核酸置換をもつものが挙げられる。

【0207】

完全にマッチする相補的ターゲットに比較して少なくとも1／2の割合でプローブにハイブリダイズする場合、即ちマッチしないターゲット核酸の任意のものとのハイブリダイゼーションに観測されるシグナル対ノイズ比の少なくとも約5倍～10倍で完全にマッチするプローブが完全にマッチする相補的ターゲットと結合する条件下におけるプローブとターゲットのハイブリダイゼーションに比較してシグナル対ノイズ比が少なくとも1／2である場合に試験核酸はプローブ核酸に特異的にハイブリダイズすると言う。

30

【0208】

核酸は一般に溶液中で会合するときに「ハイブリダイズ」する。核酸は水素結合、溶媒排除、塩基スタッキング等の種々の十分に特性決定された物理化学的力によりハイブリダイズする。核酸ハイブリダイゼーションの詳しい手引きはTijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid Probes part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays,"* (Elsevier, New York) 及びAusubel, 前出に記載されている。HamesとHiggins (1995) *Gene Probes 1* IRL Press at Oxford University Press, Oxford, 英国 (HamesとHiggins 1) 及びHamesとHiggins (1995) *Gene Probes 2* IRL Press at Oxford University Press, Oxford, 英国 (HamesとHiggins 2) はDNAとRNA（オリゴヌクレオチドを含む）の合成、標識、検出及び定量について詳細に

40

50

記載している。

【0209】

サザン又はノーザンブロットでフィルター上で100を越える相補的残基をもつ相補的核酸のハイブリダイゼーションを行うためのストリンジェントハイブリダイゼーション条件の1例は、50%ホルマリンにヘパリン1mgを加え、42℃で一晩のハイブリダイゼーションである。ストリンジェント洗浄条件の1例は65℃、0.2×SSCで15分間である（SSC緩衝液の説明についてはSambrook, 前出参照）。多くの場合には高ストリンジェンシー洗浄の前に低ストリンジェンシー洗浄でバックグラウンドプローブシグナルを除去する。低ストリンジェンシー洗浄の1例は40℃、2×SSCで15分間である。一般に、シグナル対ノイズ比が特定ハイブリダイゼーションアッセイで非関連プローブに観測されるよりも5倍（以上）である場合に特異的ハイブリダイゼーションが検出されたとみなす。

10

【0210】

サザン及びノーザンハイブリダイゼーション等の核酸ハイブリダイゼーション実験に関して「ストリンジェントハイブリダイゼーション洗浄条件」は配列依存的であり、種々の環境パラメーターにより異なる。核酸ハイブリダイゼーションの詳しい手引きはTijssen（1993），前出やHamesとHiggins 1及び2に記載されている。ストリンジェントハイブリダイゼーション及び洗浄条件は任意試験核酸について経験により容易に決定することができる。例えば、高ストリンジェントハイブリダイゼーション及び洗浄条件を決定するには、総合選択基準に合致するまで（例えばハイブリダイゼーション又は洗浄における温度上昇、塩濃度低下、界面活性剤濃度増加及び／又はホルマリン等の有機溶媒濃度増加により）ハイブリダイゼーション及び洗浄条件を漸増させる。例えば、マッチしないターゲットとプローブのハイブリダイゼーションに観測されるシグナル対ノイズ比の少なくとも約5倍でプローブが完全にマッチする相補的ターゲットと結合するまでハイブリダイゼーション及び洗浄条件を漸増させる。

20

【0211】

「超ストリンジェント」条件は特定プローブの熱融点（ T_m ）に等しくなるように選択する。 T_m は（規定イオン強度及びpH下で）試験配列の50%が完全にマッチするプローブとハイブリダイズする温度である。本発明の趣旨では、一般に規定イオン強度及びpHで特定配列の T_m よりも約5℃低くなるように「高ストリンジェント」ハイブリダイゼーション及び洗浄条件を選択する。

30

【0212】

「超高ストリンジェンシー」ハイブリダイゼーション及び洗浄条件は完全にマッチする相補的ターゲット核酸とプローブを結合させる場合のシグナル対ノイズ比がマッチしないターゲット核酸の任意のものとのハイブリダイゼーションに観測されるシグナル対ノイズ比の少なくとも10倍になるまでハイブリダイゼーション及び洗浄条件のストリンジェンシーを増す条件である。完全にマッチする相補的ターゲットのシグナル対ノイズ比の少なくとも1/2で前記条件下でプローブとハイブリダイズする場合にターゲット核酸は超高ストリンジェンシー条件下でプローブと結合すると言う。

40

【0213】

同様に、該当ハイブリダイゼーションアッセイのハイブリダイゼーション及び／又は洗浄条件を漸増させることにより更に高いレベルのストリンジェンシーを決定することもできる。例えば、完全にマッチする相補的ターゲット核酸とプローブを結合させる場合のシグナル対ノイズ比がマッチしないターゲット核酸の任意のものとのハイブリダイゼーションに観測されるシグナル対ノイズ比の少なくとも10倍、20倍、50倍、100倍又は500倍以上になるまでハイブリダイゼーション及び洗浄条件のストリンジェンシーを増す。完全にマッチする相補的ターゲットのシグナル対ノイズ比の少なくとも1/2で前記条件下でプローブとハイブリダイズする場合にターゲット核酸は超々高ストリンジェンシー条件下でプローブと結合すると言う。

50

【0214】

ストリンジェント条件下で相互にハイブリダイズしない核酸でも、これらの核酸によりコードされるポリペプチドが実質的に同一である場合には実質的に同一である。これは、例えば遺伝コードに許容される最大コドン縮重を使用して核酸のコピーを作製する場合に該当する。

【0215】

ユニークサブ配列

1 側面では、本発明は本明細書に開示するO-tRNA及びO-RSの配列から選択される核酸にユニークサブ配列を含む核酸を提供する。ユニークサブ配列は任意公知O-tRNA及びO-RS核酸配列に対応する核酸と比較してユニークである。例えばデフォルトパラメーターに設定したBLASTを使用してアラインメントを実施することができる。任意ユニークサブ配列は例えば本発明の核酸を同定するためのプローブとして有用である。

10

【0216】

同様に、本発明は本明細書に開示するO-RSの配列から選択されるポリペプチド中にユニークサブ配列を含むポリペプチドを含む。この場合には、ユニークサブ配列は公知ポリペプチド配列の任意のものに対応するポリペプチドと比較してユニークである。

【0217】

本発明はO-RSの配列から選択されるポリペプチド中のユニークサブ配列をコードするユニークコーディングオリゴヌクレオチドとストリンジェント条件下でハイブリダイズするターゲット核酸も提供し、この場合、ユニークサブ配列は対照ポリペプチド（例えば本発明のシンテターゼを得るために例えば突然変異させた親配列）の任意のものに対応するポリペプチドと比較してユニークである。ユニーク配列は上記のように決定する。

20

【0218】

配列比較、一致度及び相同道

2種以上の核酸又はポリペプチド配列に関して「一致」又は「一致度」百分率なる用語は2種以上の配列又はサブ配列を最大限に対応するように対比及び整列させ、以下に記載する配列比較アルゴリズム（又は当業者に入手可能な他のアルゴリズム）の1種を使用するか又は目視により測定した場合に相互に同一であるか又は同一のアミノ酸残基もしくはヌクレオチドの百分率が特定値であることを意味する。

【0219】

2種以上の核酸又はポリペプチド（例えばO-tRNAもしくはO-RS又はO-RSのアミノ酸配列をコードするDNA）に関して「実質的に一致」なる用語は2種以上の配列又はサブ配列を最大限に対応するように対比及び整列させ、配列比較アルゴリズムを使用するか又は目視により測定した場合にヌクレオチド又はアミノ酸残基一致度が少なくとも約60%、好ましくは80%、最も好ましくは90~95%であることを意味する。このような「実質的に一致」する配列は一般に実際の起源が記載されていなくても「相同」とあるとみなす。少なくとも約50残基長の配列の領域にわたって「実質的一致」が存在することが好ましく、少なくとも約100残基長の配列の領域がより好ましく、少なくとも約150残基又は比較する2配列の全長にわたって配列が実質的に一致していることが最も好ましい。

30

40

【0220】

配列比較及び相同性決定には、一般に一方の配列を参照配列としてこれに試験配列を比較する。配列比較アルゴリズムを使用する場合には、試験配列と参照配列をコンピューターに入力し、必要に応じてサブ配列座標を指定し、配列アルゴリズムプログラムパラメーターを指定する。こうすると、配列比較アルゴリズムは指定プログラムパラメーターに基づいて参照配列に対して試験配列の配列一致度百分率を計算する。

【0221】

比較のための最適な配列アラインメントは例えばSmith & Waterman, Adv. Appl. Math. 2: 482 (1981)の局所相同性アルゴリズム、Needleman & Wunsch, J. Mol. Biol. 48: 443 (1970)の

50

相同性アラインメントアルゴリズム、Pearson & Lipman, Proc. Nat'l Acad. Sci. USA 85:2444 (1988) の類似性探索法、これらのアルゴリズムのコンピュータソフトウェア (Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI の GAP、BESTFIT、FASTA 及び TFASTA)、又は目視 (一般に Ausubel ら, 後出参照) により実施することができる。

【0222】

配列一致度及び配列類似度百分率を決定するのに適したアルゴリズムの1例は Altschul ら, J. Mol. Biol. 215:403-410 (1990) に記載されている BLAST アルゴリズムである。BLAST 分析を実施するためのソフトウェアは National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/) から公に入手可能である。このアルゴリズムはデータベース配列中の同一長さの単語と整列した場合に所定の正の閾値スコア T と一致するか又はこれを満足するクエリー配列中の長さ W の短い単語を識別することによりまず高スコア配列対 (HSP) を識別する。 T を隣接単語スコア閾値と言う (Altschul ら, 前出)。これらの初期隣接単語ヒットをシードとして検索を開始し、これらの単語を含むもっと長い HSP を探索する。次に、累積アラインメントスコアを増加できる限り、単語ヒットを各配列に沿って両方向に延長する。ヌクレオチド配列の場合にはパラメーター M (1 対のマッチ残基のリワードスコア、常に > 0) 及び N (ミスマッチ残基のペナルティスコア、常に < 0) を使用して累積スコアを計算する。アミノ酸配列の場合には、スコアリングマトリクスを使用して累積スコアを計算する。累積アラインメントスコアがその最大到達値から量 X だけ低下するか、累積スコアが 1 カ所以上の負スコア残基アラインメントの累積によりゼロ以下になるか、又はどちらかの配列の末端に達したら各方向の単語ヒットの延長を停止する。BLAST アルゴリズムパラメーター W 、 T 及び X はアラインメントの感度と速度を決定する。BLASTIN プログラム (ヌクレオチド配列用) は語長 (W) 11、期待値 (E) 10、カットオフ 100、 $M = 5$ 、 $N = 4$ 、及び両鎖の比較をデフォルトとして使用する。アミノ酸配列用として、BLASTP プログラムは語長 (W) 3、期待値 (E) 10、及び BLOSUM62 スコアリングマトリクスをデフォルトとして使用する (Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915 参照)。

【0223】

配列一致度百分率の計算に加え、BLAST アルゴリズムは 2 配列間の類似性の統計分析も実施する (例えば Karlin & Altschul, Proc. Nat'l Acad. Sci. USA 90:5873-5877 (1993) 参照)。BLAST アルゴリズムにより提供される類似性の 1 尺度は 2 種のヌクレオチド又はアミノ酸配列間に偶然にマッチが起こる確率を示す最小合計確率 ($P(N)$) である。例えば、試験核酸を参照核酸に比較した場合の最小合計確率が約 0.1 未満、より好ましくは約 0.01 未満、最も好ましくは約 0.001 未満である場合に核酸は参照核酸に類似しているとみなす。

【0224】

免疫反応性によるポリペプチドの特性決定

本発明のポリペプチドは (例えば本発明の翻訳系で合成されるタンパク質の場合には非天然アミノ酸を含み、例えば本発明の新規シンテターゼの場合には標準アミノ酸の新規配列を含む) 種々の新規ポリペプチドを提供するので、ポリペプチドは例えばイムノアッセイで認識することができる新規構造特徴も提供する。本発明のポリペプチドに特異的に結合する抗血清の作製とこのような抗血清と結合するポリペプチドも本発明の特徴の 1 つである。

【0225】

例えば、本発明は配列番号 35 ~ 66 の 1 種以上から選択されるアミノ酸配列を含む免疫原に対して作製した抗体又は抗血清に特異的に結合するか又は特異的に免疫反応性である

10

20

30

40

50

シンテターゼタンパク質を含む。他の相同体との交差反応性をなくすために、野生型 *Methanococcus jannaschii* (*M. jannaschii*) チロシルシンテターゼ (TyrRS) 等の入手可能なシンテターゼを用いて抗体又は抗血清をサブトラクションする。野生型 *Methanococcus jannaschii* (*M. jannaschii*) チロシルシンテターゼ (TyrRS) が核酸に対応する場合には、核酸によりコードされるポリペプチドを作製し、抗体／抗血清サブトラクション目的に使用する。

【0226】

典型的なフォーマットの1例では、イムノアッセイは配列番号35～66の1種以上に対応する配列の1種以上又はその実質的サブ配列（即ち記載する全長配列の少なくとも約30%）を含む1種以上のポリペプチドに対して作製したポリクローナル抗血清を使用する。配列番号35～66に由来する潜在的ポリペプチド免疫原群を以下の文中では「免疫原性ポリペプチド」と総称する。得られた抗血清を場合により対照シンテターゼ相同体に対する交差反応性が低くなるように選択し、ポリクローナル抗血清をイムノアッセイで使用する前に例えば対照シンテターゼ相同体の1種以上に免疫吸着させることによりこのような交差反応性を除去する。

【0227】

イムノアッセイで使用する抗血清を作製するためには、免疫原性ポリペプチドの1種以上を本明細書に記載するように作製し、精製する。例えば、組換えタンパク質を組換え細胞で生産することができる。標準マウス免疫プロトコールに従って近交系マウス（マウスの仮想遺伝子一致により結果の再現性が高いのでこのアッセイで使用）に標準アジュバント（例えばフロイントアジュバント）と共に免疫原性タンパク質を免疫する（抗体作製、イムノアッセイフォーマット及び特異的免疫反応性を測定するために使用可能な条件の標準解説については、例えばHarlowとLane（1988）*Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York参照。本明細書に示した他の抗体関連文献もこの場合の免疫反応性によるポリペプチドの特性決定に適用することができる）。あるいは、本明細書に開示する配列から誘導される1種以上の合成又は組換えポリペプチドをキャリアタンパク質に結合させて免疫原として使用してもよい。

【0228】

ポリクローナル血清を収集し、イムノアッセイ（例えば固体支持体に固定化した免疫原性タンパク質の1種以上を使用する固相イムノアッセイ）で免疫原性ポリペプチドに対する力価を測定する。 10^6 以上の力価をもつポリクローナル抗血清を選択し、プールし、対照シンテターゼポリペプチドでサブトラクションし、高力価ポリクローナル抗血清サブトラクションプールを作製する。

【0229】

高力価ポリクローナル抗血清サブトラクションプールを比較イムノアッセイで対照相同体に対する交差反応性について試験する。この比較アッセイでは、高力価ポリクローナル抗血清が免疫原性シンテターゼと結合する場合のシグナル対ノイズ比が対照シンテターゼ相同体と結合する場合の少なくとも約5～10倍となるように、サブトラクション高力価ポリクローナル抗血清に差別的な結合条件を決定する。即ち、アルブミン又は脱脂粉乳等の非特異的競合剤を加えるか及び／又は塩条件、温度及び／又はその他を調節することにより結合反応のストリンジェンシーを調整する。これらの結合条件は、試験ポリペプチド（免疫原性ポリペプチド及び／又は対照ポリペプチドに比較するポリペプチド）がサブトラクションポリクローナル抗血清プールに特異的に結合するか否かを調べる後期アッセイで使用する。特に、差別的結合条件下で対照シンテターゼ相同体の少なくとも2～5倍のシグナル対ノイズ比と、免疫原性ポリペプチドの少なくとも約1.5倍のシグナル対ノイズ比を示す試験ポリペプチドは公知シンテターゼに比較して免疫原ポリペプチドと実質的な構造類似性をもつので、本発明のポリペプチドである。

【0230】

別の例では、試験ポリペプチドの検出に競合結合フォーマットのイムノアッセイを使用する。例えば、自明の通り、対照ポリペプチドに免疫吸着させることにより抗血清混合物プールから交差反応性抗体を除去する。次に、免疫原性ポリペプチドを固体支持体に固定化し、支持体をサブトラクション抗血清プールに暴露する。試験タンパク質をアッセイに加え、サブトラクション抗血清プールとの結合を競合させる。試験タンパク質が固定化タンパク質に対してサブトラクション抗血清プールとの結合を競合する能力と、アッセイに加えた免疫原性ポリペプチドが結合を競合する能力とを比較する（免疫原性ポリペプチドは抗血清プールとの結合を固定化免疫原性ポリペプチドと有効に競合する）。標準計算を使用して試験タンパク質の交差反応性百分率を計算する。

【0231】

平行アッセイで場合により対照タンパク質がサブトラクション抗血清プールとの結合を競合する能力を免疫原性ポリペプチドが抗血清との結合を競合する能力と比較測定する。この場合も、標準計算を使用して対照ポリペプチドの交差反応性百分率を計算する。試験ポリペプチドの交差反応性百分率が対照ポリペプチドの少なくとも5～10倍である場合又は試験ポリペプチドの結合が免疫原性ポリペプチドの結合とほぼ同一範囲である場合に、試験ポリペプチドはサブトラクション抗血清プールに特異的に結合すると言う。

【0232】

一般に、本明細書に記載する競合的結合イムノアッセイでは任意試験ポリペプチドを免疫原性及び／又は対照ポリペプチドに比較するために免疫吸着抗血清プールを使用することができる。この比較を行うためには、免疫原性、試験及び対照ポリペプチドを各々広い濃度範囲でアッセイし、サブトラクション抗血清と例えば固定化した対照、試験又は免疫原性タンパク質との結合の50%を阻害するために必要な各ポリペプチドの量を標準技術により決定する。競合アッセイで結合に必要な試験ポリペプチドの量が免疫原性ポリペプチドの必要量の2倍未満であり、対照ポリペプチドの少なくとも約5～10倍である場合には試験ポリペプチドは免疫原性タンパク質に対して作製した抗体に特異的に結合すると言う。

【0233】

特異性の付加試験として、得られる免疫原性ポリペプチドサブトラクション抗血清プールと免疫吸着に使用する免疫原性ポリペプチドの結合が殆ど又は全く検出できなくなるまで場合により抗血清プールを（対照ポリペプチドではなく）免疫原性ポリペプチドに完全に免疫吸着させる。この完全に免疫吸着した抗血清を次に試験ポリペプチドとの反応性について試験する。反応性が殆ど又は全く観察されない場合（即ち完全に免疫吸着した抗血清と免疫原性ポリペプチドの結合に観察されるシグナル対ノイズ比の2倍以下）には、試験ポリペプチドは免疫原性タンパク質により誘導される抗血清に特異的に結合する。

【0234】

一般分子生物学技術

分子生物学技術について記載している一般教科書としてはBergerとKimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrookら, Molecular Cloning—A Laboratory Manual (第2版), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 (「Sambrook」); 及び Current Protocols in Molecular Biology, F. M. Ausubelら, 編, Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1999年補遺) (「Ausubel」) が挙げられる。これらの教科書は突然変異誘発、ベクターの使用、プロモーター及び他の多くの関連事項について記載しており、例えば非天然アミノ酸、直交tRNA、直交シンテターゼ及びその対を含むタン

10

20

30

40

50

パク質を生産するためのセクターコドンを含む遺伝子の作製についても記載している。

【0235】

例えば非天然アミノ酸をコードするセクターコドンをタンパク質に挿入するために、本発明では各種突然変異誘発を使用する。これらの方法としては部位特異的、ランダム点突然変異誘発、相同組換え（DNAシャフリング）、ウラシル含有鋳型を使用する突然変異誘発、オリゴヌクレオチド特異的突然変異誘発、ホスホロチオエート修飾DNA突然変異誘発、ギャップデュプレクスDNAを使用する突然変異誘発等が挙げられるが、これらに限定されない。他の利用可能な方法としては点ミスマッチ修復、修復欠損宿主株を使用する突然変異誘発、制限－選択及び制限－精製、欠失突然変異誘発、完全遺伝子合成による突然変異誘発、2本鎖切断修復等が挙げられる。例えばキメラ構築物を使用する突然変異誘発も本発明に含まれる。1態様では、天然分子又は改変もしくは突然変異させた天然分子の公知情報（例えば配列、配列比較、物性、結晶構造等）により突然変異誘発を実施することができる。

10

【0236】

本明細書に記載する上記教科書と実施例はこれらの手順について記載している。以下の刊行物とその引用文献にも情報が記載されている。Lingら, Approaches to DNA mutagenesis: an overview, Anal Biochem. 254 (2): 157-178 (1997); Daleら, Oligonucleotide-directed random mutagenesis using phosphorothioate method, Methods Mol. Bio 20 57: 369-374 (1996); Smith, In vivo mutagenesis, Ann. Rev. Genet. 19: 423-462 (1985); Botstein & Shortle, Strategies and applications of in vitro mutagenesis, Science 229: 1193-1201 (1985); Carter, Site-directed mutagenesis, Biochem. J. 237: 1-7 (1986); Kunkel, The efficiency of oligonucleotide directed mutagenesis, in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D. M. J. 編, Springer Verlag, Berlin) (1987); Kunklel, R 30 apid and efficient site-specific mutagenesis without phenotypic selection, Proc. Natl. Acad. Sci. USA 82: 488-492 (1985); Kunkelら, Rapid and efficient site-specific mutagenesis without phenotypic selection, Methods in Enzymol. 154, 367-382 (1987); Bassら, Mutant Trp repressors with new DNA-binding specificities, Science 242: 240-245 (1988); Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Zol 40 ller & Smith, Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment, Nucleic Acids Res. 109: 6487-6500 (1982); Zoller & Smith, Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, Methods in Enzymol. 100: 468-500 (1983); Zoller & Smith, Oligonucleotide-directed mutagenesis: a simple met 50

hod using two oligonucleotide primers and a single-stranded DNA template, Methods in Enzymol. 154:329-350 (1987); Taylor⁵, The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA, Nucl. Acids Res. 13:8749-8764 (1985); Taylor⁵, The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA, Nucl. Acids Res. 13:8765-8787 (1985) 10
); Nakamaye² & Eckstein, Inhibition of restriction endonuclease NciI cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis, Nucl. Acids Res. 14:9679-9698 (1986); Sayers⁵, Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, Nucl. Acids Res. 16:791-802 (1988); Sayers⁵, Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide, (1988) Nucl. Acids Res. 16:803-814; Kramer⁵, The gapped duplex DNA approach to oligonucleotide-directed mutation construction, Nucl. Acids Res. 12:9441-9456 (1984); Kramer & Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, Methods in Enzymol. 154:350-367 (1987); Kramer⁵, Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, Nucl. Acids Res. 16:7207 (1988); Fritz⁵, Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro, Nucl. Acids Res. 16:6987-6999 (1988); Kramer⁵, Point Mismatch Repair, Cell 38:879-887 (1984); Carter⁵, Improved oligonucleotide site-directed mutagenesis using M13 vectors, Nucl. Acid s Res. 13:4431-4443 (1985); Carter, Improved oligonucleotide-directed mutagenesis using M13 vectors, Methods in Enzymol. 154:382-403 (1987); Eghtedarzadeh & Henikoff, Use of oligonucleotides to generate large deletions, Nucl. Acids Res. 14:5115 (1986); Wells⁵, Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin, Phil. Trans. R. Soc. Lond. A 317:415-423 (1986); Nambiar⁵, Total synthesis 50

and cloning of a gene coding for the ribonuclease S protein, Science 223:1299-1301 (1984); Sakamar & Khorana, Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin), Nucl. Acids Res. 14:6361-6372 (1988); Wellsら, Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites, Gene 34:315-323 (1985); Grundstromら, Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis, Nucl. Acids Res. 13:3305-3316 (1985); Mandecki, Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis, Proc. Natl. Acad. Sci. USA, 83:7177-7181 (1986); Arnold, Protein engineering for unusual environments, Current Opinion in Biotechnology 4:450-451 (1993); Sieberら, Nature Biotechnology, 19:456-460 (2001); W. P. C. Stemmer, Nature 370, 389-91 (1994); 及び I. A. Lorimer, I. Pastan, Nucleic Acids Res. 23, 3067-8 (1995)。上記方法の多くに関する更に詳細な説明は Methods in Enzymology Volume 154 に記載されており、この文献には各種突然変異誘発法に伴うトラブルシューティング問題の有用な解決方法も記載されている。

【0237】

本発明は直交 tRNA/R S 対を介する非天然アミノ酸の *in vivo* 組込みに用いる宿主細胞及び生物にも関する。例えばクローニングベクター又は発現ベクターとすることができる本発明のベクターで宿主細胞を遺伝子組換え（例えば形質転換、形質導入又はトランスフェクション）する。ベクターは例えばプラスミド、細菌、ウイルス、裸のポリヌクレオチド又はポリヌクレオチドコンジュゲートの形態とすることができる。ベクターはエレクトロポレーション（Fromら, Proc. Natl. Acad. Sci. USA 82, 5824 (1985)）、ウイルスベクターによる感染、核酸を小ビーズもしくは粒子のマトリックスに埋込むか又は表面に付着させて小粒子形態で高速射入する（Kleinら, Nature 327, 70-73 (1987)）等の標準方法により細胞及び/又は微生物に導入する。

【0238】

組換え宿主細胞は例えばスクリーニング段階、プロモーター活性化又は形質転換細胞選択等の操作に合うように適宜改変した慣用栄養培地で培養することができる。これらの細胞を場合によりトランスジェニック生物で培養することができる。

【0239】

ターゲット核酸を宿主細胞に導入する方法としては数種の周知方法が入手可能であり、本発明ではその任意のものを使用することができる。これらの方法としては、DNAを含む細菌プロトプラストとレシピエント細胞の融合、エレクトロポレーション、遺伝子銃及びウイルスベクターによる感染（下記に詳述）等が挙げられる。本発明のDNA構築物を含むプラスミドの数を増幅するためには細菌細胞を使用することができる。対数期まで細菌細胞を増殖させると、細菌内のプラスミドを当分野で公知の各種方法により分離することができる（例えば Sambrook 参照）。更に、細菌からプラスミドを精製するために

多数のキットが市販されている（例えばEasyPrep（登録商標）、FlexiPrep（登録商標）（いずれもPharmacia Biotech）；StrataClean（登録商標）（Stratagene）；及びQIAprep（登録商標）（Qiagen））。分離精製したプラスミドを更に操作して他のプラスミドを作製し、細胞をトランスフェクトするために使用するか又は生物に感染させるために関連ベクターに組込む。典型的ベクターは転写及び翻訳ターミネーターと、転写及び翻訳開始配列と、特定ターゲット核酸の発現の調節に有用なプロモーターを含む。ベクターは場合により少なくとも1個の独立ターミネーター配列と、真核生物又は原核生物又は両者（例えばシャトルベクター）でカセットの複製を可能にする配列と、原核系と真核系の両者の選択マーカーを含む包括的発現カセットを含む。ベクターは原核生物、真核生物、又は好ましくは両者での複製と組込みに適している。Giliman & Smith, Gene 8:81 (1979)；Robertsら, Nature, 328:731 (1987)；Schneider, B. ら, Protein Expr. Purif. 6435:10 (1995)；Ausubel, Sambrook, Berger（いずれも前出）参照。クローニングに有用な細菌とバクテリオファージのカタログは例えばATCCから入手でき、例えばATCCから刊行されたThe ATCC Catalogue of Bacteria and Bacteriophage (1992) Ghernaら（編）が挙げられる。その他のシーケンシング、クローニング及び分子生物学の他の側面の基本手順と基礎理論事項もWatsonら（1992）Recombinant DNA 第2版, Scientific American Books, NYに記載されている。

10

20

【0240】

（例えば後期核酸分離のための）例えば細胞分離及び培養に関する他の有用な文献としてはFreshney (1994) Culture of Animal Cells, a Manual of Basic Technique, 第3版, Wiley-Liss, New Yorkとその引用文献；Payneら (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc, New York, NY；GamborgとPhillips（編）(1995) Plant Cell, Tissue and Organ Culture；Fundamental Method Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York；及びAtlasとParks（編）The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FLが挙げられる。

30

【0241】

更に、The Midland Certified Reagent Company (mcrc@oligos.com)、The Great American Gene Company (www.genco.com)、ExpressGen Inc. (www.expressgen.com)、Operon Technologies Inc. (Alameda, CA)、その他多数の各種販売会社からほぼ任意核酸（及び標準又は標準外を問わずほぼ任意標識核酸）をオーダーメイド又は標準注文することができる。

40

【0242】

医薬組成物

（例えば1種以上の非天然アミノ酸を含む）本発明のタンパク質（例えばポリペプチド、ペプチド等）は場合により例えば適当な医薬キャリアと共に治療用途に使用される。このような組成物は例えば治療上有効な量の化合物と医薬的に許容可能なキャリア又は賦形剤を含む。このようなキャリア又は賦形剤としては食塩水、緩衝食塩水、デキストロース、水、グリセロール、エタノール及び／又はその組合せが挙げられるが、これらに限定されない。製剤は投与方法に合わせて製造される。一般に、タンパク質の投与方法は当分野で周知であり、本発明のポリペプチドの投与に適用することができる。

50

【0243】

1種以上の本発明のポリペプチドを含む治療用組成物は場合により1種以上の適当な *in vitro* 及び／又は *in vivo* 動物疾患モデルで試験し、当分野で周知の方法により効力、組織代謝を確認し、用量を推定する。特に、本明細書に記載する非天然アミノ酸の活性、安定性又は他の適当な尺度を天然アミノ酸相同体と比較（例えば1種以上の非天然アミノ酸を組込むように改変したEPOと天然アミノ酸EPOを比較）する試験即ち関連アッセイによりまず用量を決定することができる。

【0244】

投与は分子を最終的に血液又は組織細胞と接触させるために通常使用されている任意経路で行う。本発明の非天然アミノ酸ポリペプチドは場合により1種以上の医薬的に許容可能なキャリアと共に適当な任意方法で投与される。本発明の関連でこのようなポリペプチドを患者に投与するのに適した方法は入手可能であり、特定組成物を投与するために2種以上の経路を使用することもできるが、特定経路のほうが別の経路よりも迅速で有効な作用又は反応を得られることが多い。

10

【0245】

医薬的に許容可能なキャリアは投与する特定組成物や、組成物を投与するために使用する特定方法によっても異なる。従って、本発明の医薬組成物の利用可能な製剤は多様である。

【0246】

ポリペプチド組成物は多数の経路で投与することができ、例えば経口、静脈内、腹膜組織内、筋肉内、経皮、皮下、局所、舌下又は直腸手段が挙げられるが、これらに限定されない。非天然アミノ酸ポリペプチド組成物はリポソームにより投与することもできる。このような投与経路及び適切な製剤は一般に当業者に公知である。

20

【0247】

非天然アミノ酸ポリペプチドは単独又は他の適当な成分と共にエアゾール製剤に処方（即ち「噴霧」）して吸入投与することもできる。エアゾール製剤はジクロロジフルオロメタン、プロパン、窒素等の許容可能な加圧噴射剤に加えることができる。

【0248】

例えば関節内、静脈内、筋肉内、皮内、腹膜組織内及び皮下経路等の非経口投与に適した製剤としては、酸化防止剤、緩衝液、静菌剤、及び製剤を所期レシピエントの血液に等張にする溶質を加えることができる水性及び非水性等張滅菌注射溶液剤や、懸濁剤、溶解補助剤、増粘剤、安定剤及び防腐剤を加えることができる水性及び非水性滅菌懸濁液が挙げられる。アンプルやバイアル等の単位用量又は多重用量密閉容器に収容してパッケージ核酸製剤とすることもできる。

30

【0249】

非経口投与と静脈内投与が好ましい投与方法である。特に、天然アミノ酸相同体治療で既に使用されている投与経路（例えばEPO、GCSF、GMCSF、IFN、インターロイキン、抗体及び／又は他の医薬的送達タンパク質に一般に使用される投与経路）と現行製剤が本発明の非天然アミノ酸の好ましい投与経路と製剤である。

【0250】

本発明の関連で患者に投与する用量は患者に有益な治療応答を経時的に生じるために十分な量又は例えば用途に応じて病原体による感染を阻害するか又は他の適当な活性を生じるために十分な量とする。用量は特定ベクター又は製剤の効力、使用する非天然アミノ酸ポリペプチドの活性、安定性又は血清半減期、患者の症状、更には治療する患者の体重又は体表面積により異なる。用量は特定患者における特定ベクター、製剤等の投与に伴う副作用の存在、性質及び程度によっても異なる。

40

【0251】

疾病（例えば癌、遺伝病、糖尿病、エイズ等）の治療又は予防に投与するベクター又は製剤の有効量を決定する際には、医師は循環血漿値、製剤毒性、疾病の進行及び／又は該当する場合には抗非天然アミノ酸ポリペプチド抗体の生産を考慮する。

50

【0252】

例えば体重70kgの患者の用量は一般に現在使用されている治療用タンパク質の用量と同等の範囲であり、該当組成物の活性又は血清半減期の変化に合わせて調節する。本発明のベクターは抗体投与、ワクチン投与、細胞毒性剤、天然アミノ酸ポリペプチド、核酸、ヌクレオチド類似体、生体応答調節剤の投与等の任意公知慣用治療による治療条件を補うことができる。

【0253】

投与に当たり、本発明の製剤は該当製剤のLD50及び／又は例えば患者の体重と総合健康状態に応じた各種濃度の非天然アミノ酸の副作用の観察により決定される頻度で投与する。一度に投与してもよいし、分けて投与してもよい。

10

【0254】

製剤を注入した患者が発熱、悪寒又は筋肉痛を生じる場合には、適当な用量のアスピリン、イブプロフェン、アセトアミノフェン又は他の鎮痛／解熱剤を投与する。発熱、筋肉痛及び悪寒等の注入反応を生じる患者には、次回注入の30分前にアスピリン、アセトアミノフェン又は例えばジフェニルヒドラミンを前投薬する。解熱剤や抗ヒスタミン剤がすぐに効かない重度悪寒及び筋肉痛にはメペリジンを使用する。反応の重度に応じて細胞注入を遅らせるか又は中断する。

【実施例1】

【0255】

実施例1：O-メチル-L-チロシンのin vivo組込み

20

異種間アミノアシル化が非効率的な場合には、別の生物からの対を導入することにより大腸菌で直交tRNA／シンテターゼ対を作製することができ(Y. Kwok, J. T. Wong, Can. J. Biochem. 58, 213-8 (1980))、アンチコドンループはシンテターゼ認識の主要決定因子ではない。このような対の1候補はtRNAアイデンティティーエレメントが大腸菌tRNA^{Tyr}と異なり、チロシルシンテターゼ(TyrRS)がアンチコドンループ結合ドメインを欠失する始原菌であるMethanococcus jannaschii (M. jannaschii)のチロシルtRNA／シンテターゼ対である(B. A. Steer, P. Schimmel, J. Biol. Chem. 274, 35601-6 (1999))。更に、M. jannaschii TyrRSは編集メカニズムをもたない(H. Jakubowski, E. Goldman, Microbiol. Rev. 56, 412-29 (1992))ので、tRNAにライゲートした非天然アミノ酸を校正しないと思われる。

30

【0256】

M. jannaschii tRNA^{Tyr}に由来するアンバーサプレッサーtRNAは大腸菌シンテターゼにより効率的にアミノアシル化されないが、大腸菌でタンパク質翻訳に効率的に機能することが示されている(L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, J. Am. Chem. Soc. 122, 5010-1 (2000))。更に、M. jannaschii TyrRSは大腸菌tRNAに直交性である(B. A. Steer, P. Schimmel, J. Biol. Chem. 274, 35601-6 (1999))が、それ自体のサプレッサーtRNA^{CUA}^{Tyr}を効率的にアミノアシル化する(L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, J. Am. Chem. Soc. 122, 5010-1 (2000))。従って、M. jannaschii tRNA^{CUA}^{Tyr}／TyrRSは直交対として機能し、大腸菌でアンバーコドンUAGに応答してチロシンを効率的に組込むことができる。

40

【0257】

大腸菌シンテターゼによるこの直交tRNAの認識を更に弱めるために、突然変異誘発及び選択スキームを実施した。更に詳細については、本願と同時に出願し、その開示内容全体を本明細書に組込む米国特許出願番号第_____号(発明の名称"Methods and Compositions for the Production of Or

50

thogonal tRNA-tRNA Synthetase Pairs") (代理人整理番号54-000130US, Schultzら)を参照されたい。

【0258】

要約すると、M. jannaschii TyrRSと直接相互作用しないM. jannaschii tRNA_{CUA}^{Tyr} CUA Tyrの11個のヌクレオチドをランダムに突然変異させてサプレッサーtRNAライブラリーを得た。このtRNAライブラリーをネガティブ選択にかけて大腸菌シンテターゼによりアミノアシル化されるtRNAを除去した後、ポジティブ選択にかけ、M. jannaschii TyrRSにより効率的にアミノアシル化されるtRNAを選択した。得られたサプレッサーtRNAの直交性をプラスミドpBLAMに担持したTEM-1βラクタマーゼ遺伝子の非必須位置 (Ala184) のアンバー終止コドンの抑圧に基づいてin vivo相補アッセイにより試験した。形質転換したサプレッサーtRNAが内在大腸菌シンテターゼによりアミノアシル化されると、アンピシリンの存在下で細胞増殖する。M. jannaschii tRNA_{CUA}^{Tyr} とpBLAMで形質転換した大腸菌は55.5 μg/mLアンピシリンの存在下に生存する。ライブラリーから選択した最良突然変異体サプレッサーtRNA (mtRNA_{CUA}^{Tyr}) が発現された場合には、細胞は僅か12.4 μg/mLのアンピシリン濃度でも生存した。突然変異体サプレッサーmtRNA_{CUA}^{Tyr} はヌクレオチド置換C17A、U17aG、U20C、G37A及びU47Gを含んでいた。比較のために、pBLAMのみを導入した細胞は(サプレッサーtRNAの不在下で)アンピシリン濃度9.7 μg/mLで生存する。M. jannaschii TyrRSをこのmtRNA_{CUA}^{Tyr} と同時発現させると、細胞はアンピシリン濃度436 μg/mLで生存する。従って、mtRNA_{CUA}^{Tyr} はM. jannaschii tRNA_{CUA}^{Tyr} よりも弱い内在シンテターゼの基質であるので、M. jannaschii TyrRSにより効率的にアミノアシル化される。

【0259】

mtRNA_{CUA}^{Tyr} に所望非天然アミノ酸を負荷するように直交M. jannaschii TyrRSのアミノ酸特異性を改変するために、TyrRS突然変異体のライブラリーを作製し、スクリーニングした。Bacillus stearothermophilusに由来する相同TyrRSの結晶構造(P. Brick, T. N. Bhat, D. M. Blow, J. Mol. Biol. 208, 83-98 (1988))に基づき、結合チロシンのアリール環のパラ位から6.5 Å以内にあるM. jannaschii TyrRSの活性部位の5個の残基(Tyr32、Glu107、Asp158、Ile159及びLeu162)を突然変異させた。突然変異体M. jannaschii TyrRS (mutTyrRS、O-メチル-L-チロシンによるアミノアシル化の場合)からの対応する残基はTyr³² (Tyr³⁴)、Glu¹⁰⁷ (Asn¹²³)、Asp¹⁵⁸ (Asp¹⁷⁶)、Ile¹⁵⁹ (Phe¹⁷⁷)及びLeu¹⁶² (Leu¹⁸⁰)である(括弧内はB. stearothermophilus TyrRS残基を表す)。

【0260】

以下に更に詳述するように、これらの残基をまず全てアラニンに突然変異させ、mtRNA_{CUA}^{Tyr} にチロシンを負荷することができないAla5TyrRSを作製した。この突然変異体Ala5TyrRSを鋳型として使用し、オリゴヌクレオチドをプライマーとしてPCR突然変異誘発により5個の残基をランダム突然変異させたTyrRS突然変異体のライブラリーを作製した。

【0261】

カナマイシン耐性を付与したpBR322由来プラスミドであるプラスミドpBK-JYRSで大腸菌GlnRSプロモーター及びターミネーターの制御下にM. jannaschii TyrRS遺伝子を発現させた。残基Tyr32、Glu107、Asp158、Ile159及びLeu162を部位特異的突然変異誘発によりAlaで置換し、プラスミドpBK-JYA5を得た。オリゴヌクレオチドLW157 5'-GGAAATTC CATATGGACGAATTGTGA AATG-3'、LW164 5'-GTATTT

T A C C A C T T G G T T C A A A A C C T A T M N N A G C A G A T T T T T C A T C
 T T T T T T T C A T C T T T T T T A A A A C - 3'、LW159 5' - T A G G
 T T T T G A A C C A A G T G G T A A A A T A C - 3'、LW165 5' - C A T T
 C A G T G T A T A A T C C T T A T C A A G C T G G A A M N N A C T T C C A T A A
 A C A T A T T T T G C C T T T A A C - 3'、LW161 5' - T C C A G C T T G
 A T A A G G A T T A T A C A C T G A A T G - 3'、LW167 5' - C A T C C C
 T C C A A C T G C A A C A T C A A C G C C M N N A T A A T G M N N M N N A T T A
 A C C T G C A T T A T T G G A T A G A T A A C - 3'、LW163 5' - G C G T
 T G A T G T T G C A G T T G G A G G G A T G - 3'、及びLW105 5' - A A A
 C T G C A G T T A T A A T C T C T T T C T A A T T G G C T C - 3' とNNK (N = 10
 A + T + G + C、K = G + T、及びM = C + A (Operon, Alameda, CA)
 を突然変異部位に使用してAla5TyrRS突然変異体(pBK-JYA5)のPCR
 増幅を行い、NdeI-PstIで消化したpBK-JYA5に再ライゲートしてTyr
 RSライブラリーを得た。ライゲートしたベクターを大腸菌DH10Bコンピテント細胞
 に形質転換し、 1.6×10^9 コロニー形成単位(cfu)のライブラリーを得た。40
 個のランダムに抽出したコロニーからのTyrRS遺伝子を配列決定した処、ランダムに
 突然変異させたNNK位置に塩基変異はなく、他の予想外の突然変異もないことが確認さ
 れた。ライブラリーをマキシプレップにより増幅し、スーパーコイルDNAを使用して選
 択株pYC-J17を形質転換した。

【0262】

次に、クロラムフェニコールアセチルトランスフェラーゼ(CAT)遺伝子の非必須位置
 (Asp112)のアンバーコドンの抑圧に基づくポジティブ選択を行った(M. Pas
 trnak, T. J. Magliery, P. G. Schultz, Helvetica
 Chimica Acta. 83, 2277-86 (2000))。非天然アミノ酸を
 加えた培地で細胞を増殖させ、各種濃度のクロラムフェニコールの存在下でのその生存に
 ついて選択した。突然変異体TyrRSが直交mtRNA_{CUA}^{Tyr}に天然又は非天然アミ
 ノ酸を負荷する場合には、細胞はCATを生産し、生存する。

【0263】

次に生存細胞をクロラムフェニコールの存在下で非天然アミノ酸の不在下に増殖させた。
 非天然アミノ酸を加えたレプリカプレートから生存しない細胞即ち直交mtRNA_{CUA}^{Ty} 30
 「に非天然アミノ酸を負荷する突然変異体TyrRSをコードする細胞を分離した。突然
 変異体TyrRS遺伝子をこれらの細胞から分離し、DNAシャフリングによりin v
 itro組換えし、後期選択のために大腸菌に再形質転換した。

【0264】

アリール環のパラ位に各種官能基(アセチル、アミノ、カルボキシル、イソプロピル、メ
 チル、オーメチル及びニトロ)をもつ7種のチロシン類似体を個々に使用して以下の選択
 を行った。

【0265】

lppプロモーターとrrnCターミネーターの制御下にmtRNA_{CUA}^{Tyr}をコードす
 る遺伝子をプラスミドpACMD112TAG(そのCAT遺伝子のAsp112を置換 40
 するTAG終止コドンをつけたpACYC184プラスミド(M. Pasternak, T.
 J. Magliery, P. G. Schultz, Helvetica Chimica
 Acta. 83, 2277-86 (2000))に挿入し、プラスミドpYC-J
 17を得た。pYC-J17を導入した大腸菌DH10Bコンピテント細胞にTyrRS
 ライブラリーをコードするスーパーコイルDNAを形質転換し、元のライブラリーを完全
 にカバーする 3×10^9 cfuよりも大きい寸法のライブラリーを得た。次に、1%グリ
 セロールと0.3mMロイシンを含む最少培地(GMML)プレートに17µg/mLテ
 トラサイクリン(Tet)、25µg/mLカナマイシン(Kan)、50µg/mLク
 ロラムフェニコール(Cm)及び1mM非天然アミノ酸を加えて細胞を培養した。37℃
 で44時間培養後にオーメチルーL-チロシンを加えたプレート上のコロニーをプールし 50

、プラスミドを分離し、pYC-J17を導入した大腸菌DH10Bコンピテント細胞に再形質転換し、形質転換細胞を $50\mu\text{g}/\text{mL}$ Cm上でポジティブ選択した。コロニー(96)をプレートから個々に抽出し、液体GMM L培地 100mL で希釈し、各種濃度のCmを加えたKan/Tet GMM Lプレート2組に画線した。第1組のプレートはO-メチル-L-チロシンを加えず、Cm濃度を $10\sim 25\mu\text{g}/\text{mL}$ とし、第2組のプレートは 1mM O-メチル-L-チロシンと $50\mu\text{g}/\text{mL}$ Cmを加えた。第1組のプレートで $15\mu\text{g}/\text{mL}$ のCmの存在下に増殖しなかったコロニーのレプリカを第2組のプレートから抽出した。Ty r R S遺伝子を含むプラスミドを精製し、断片化反応の Mg^{2+} を 10mM Mn^{2+} に代えた以外はStemmerのプロトコル(W. P. C. Stemmer, Nature 370, 389-91 (1994))を使用してDNAシャフリングによりin vitro組換えした(I. A. Lorimer, I. Pastan, Nucleic Acids Res. 23, 3067-8 (1995))。次に、予め消化しておいたpBK-JYA5ベクターにライブラリーを再ライゲートし、典型寸法 $8\times 10^8\sim 3\times 10^9\text{cfu}$ の第2世代Ty r R Sライブラリーを得た。ライブラリーからランダムに選択した30個を配列決定した。DNAシャフリングにより導入された突然変異誘発率は0.35%であった。このライブラリーを次の選択のために選択株に形質転換した後、シャフリングした。ポジティブ選択と第2組プレートのCm濃度は2回目に $80\mu\text{g}/\text{mL}$ 、3回目に $120\mu\text{g}/\text{mL}$ に上げ、第1組のプレートのCm濃度は変えなかった。3回のDNAシャフリング後、コロニーは第1組のプレートではCm濃度 $20\sim 25\mu\text{g}/\text{mL}$ で増殖し始め、Ty r R S突然変異体が天然アミノ酸を基質として受容していることを示した。従って、2回のDNAシャフリング後に選択した最良クローンを詳細に特性決定した。

【0266】

2回の選択とDNAシャフリング後にクロラムフェニコール中の生存が増殖培地への 1mM O-メチル-L-チロシン添加に依存するクローン(突然変異体Ty r R S(LWJ16))を得た。O-メチル-L-チロシンの不在下では、突然変異体Ty r R S(LWJ16)を導入した細胞は1%グリセロール、0.3mMロイシンを含む最少培地(GMM L)プレートに $15\mu\text{g}/\text{mL}$ クロラムフェニコールを加えると生存できなかった。細胞は 1mM O-メチル-L-チロシンの存在下では $125\mu\text{g}/\text{mL}$ クロラムフェニコールを加えたGMM Lプレートで増殖することができた。液体GMM Lでも同様の結果が得られた。対照として、mtRNA_{CUA}^{Tyr}と不活性Ala5Ty r R Sを導入した細胞は 1mM O-メチル-L-チロシンの有無に拘わらず、使用した最低クロラムフェニコール濃度でも生存しなかった。これは、クロラムフェニコール中の細胞増殖が突然変異体Ty r R S(LWJ16)の発現に依存しており、単なるO-メチル-L-チロシンの栄養効果ではないことを示している。 1mM O-メチル-L-チロシン自体を加えても大腸菌の増殖率はさほど変わらない。

【0267】

観察される表現型がアンバー終止コドンに応答する直交mtRNA_{CUA}^{Tyr}/突然変異体Ty r R S(LWJ16)対によるO-メチル-L-チロシンの部位特異的組込みに起因することを更に立証するために、ジヒドロ葉酸レダクターゼ(DHFR)のO-メチル-L-チロシン突然変異体を作製し、特性決定した。大腸菌DHFR遺伝子の第3コドン(許容部位)をTAGに突然変異させ、突然変異体タンパク質を内在大腸菌DHFRから分離するためにC末端His₆タグを付けた。対照として、mtRNA_{CUA}^{Tyr}を野生型M. jannaschii Ty r R Sと同時発現させ、DHFR中のナンセンスコドンをチロシンで効率的に抑圧した。図2参照。液体GMM L増殖培地でmtRNA_{CUA}^{Tyr}と 1mM O-メチル-L-チロシンの存在下に突然変異体Ty r R S(LWJ16)を発現させると、全長DHFRも生産され、分離収率 $2\text{mg}/\text{L}$ でNiアフィニティークロマトグラフィーにより精製することができた。

【0268】

精製タンパク質の収率は液体GMM L培地では2YTリッチ培地の約26分の1である。

例えば、 $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ と野生型 $\text{M. jannaschii TyrRS}$ を同時発現させる場合には、 DHFR の収率は 2YT では 67mg/L 、液体 GMM L では 2.6mg/L である。

【0269】

オーメチル-L-チロシン、 $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ 又は突然変異体 TyrRS (LWJ16) の不在下では SDS -ポリアクリルアミドゲル電気泳動と銀染色による分析で DHFR は観察されなかった（デンシトメトリーによると $<0.1\%$ ）。図2参照。更にウェスタン分析の結果、 $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ 、突然変異体 TyrRS (LWJ16) 又はオーメチル-L-チロシンの不在下では DHFR が全く生産されないことが判明した。図2参照。

【0270】

TAG コドンに応答して挿入されたアミノ酸の正体は無傷のタンパク質とトリプシン消化フラグメントの両者の質量分析により確認した。無傷のタンパク質の平均分子量をエレクトロスプレーイオン化フーリエ変換イオンサイクロトロン共鳴質量スペクトロメトリー（ FT-ICR MS ）により測定した。内部標準に近接するクラスターからのモノアイソトピック質量の観測値は 18096.002Da であり、理論質量 18095.908Da から 5ppm の範囲内であり、オーメチル-L-チロシンの組込みを明らかに示している。

【0271】

この実験には C 末端 His タグを付けない DHFR 突然変異体を使用し、メトトレキサートアフィニティークロマトグラフィーにより精製した。突然変異体タンパク質ではアンバー抑圧効率を高めるために第3コドンを TAG に置換し、第4コドンを CTG から ATG に置換し、 Leu4Met 突然変異とした。

【0272】

この結果も他の内在大腸菌シンテターゼがオーメチル-L-チロシンを基質として利用しないことを示している。トリプシン消化物の液体クロマトグラフィータンデム質量分析を行い、 N 末端ペプチドの配列を確認した。タンデム MS スペクトルの1例を図3に示す。 N 末端トリプシン消化ペプチド $\text{MIY}^*\text{MIAALAVDR}$ に対応する 691.5Da の二重負荷前駆体イオンを選択し、イオントラップ質量分析計（ ITMS ）で断片化した。フラグメントイオン質量は図3に示すようにはっきりと割当られ、オーメチル-L-チロシンの部位特異的組込みを裏付けた。タンパク質質量スペクトルもトリプシン消化ペプチドマップもオーメチル-L-チロシンの代わりにチロシン又は他のアミノ酸が組込まれたことを示さず、タンパク質質量スペクトルのシグナル対ノイズ比からオーメチル-L-チロシンの最低組込み純度 95% が得られた。

【0273】

まとめると、細胞増殖、タンパク質発現及び質量分析実験の結果、 $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ / 突然変異体 TyrRS (LWJ16) 直交対は天然アミノ酸に匹敵する忠実度でアンバーコドンに応答してオーメチル-L-チロシンをタンパク質に選択的に挿入できることが判明した。

【0274】

$\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ にオーメチル-L-チロシンを負荷する突然変異体 TyrRS (LWJ16) の配列を分析した処、12箇所のヌクレオチド置換が判明し、そのうち2箇所はサイレントであった。10箇所の非サイレント突然変異の結果、野生型 TyrRS に対して以下のアミノ酸残基置換が生じた。水素が天然基質チロシンのアリール酸素原子に結合する Tyr32 は Gln に突然変異し、水素がチロシンのヒドロキシル基に結合する Asp158 は Ala に突然変異し、水素が Asp158 に結合する Glu107 は Thr に突然変異し、結合ポケットの底部に位置する Leu162 は Pro に突然変異した。相同 $\text{B. stearothermophilus TyrRS}$ のX線結晶構造によると、 Tyr32 、 Asp158 と基質チロシン間の水素結合網が失われるとチロシンと突然変異体 TyrRS (LWJ16) の結合に不利であると思われる。実際に、 $\text{B. stearothermophilus TyrRS}$ の Asp176 (M. jannaschii の As

10

20

30

40

50

p 1 5 8 に対応) の突然変異は不活性な酵素を生じる (G. D. P. Gray, H. W. Duckworth, A. R. Fersht, FEBS 318, 167-71 (1993))。同時に、Asp 1 5 8 Ala 及び Leu 1 6 2 Pro 突然変異は O-メチル-L-チロシンのメチル基を更に基質結合キャビティに延ばす疎水性ポケットを形成する。リボース又はアデニル酸のリン酸基に結合する活性部位の他の重要な触媒残基は 2 回の DN A シャフリング後も変化しなかった。これらの突然変異の詳細な分析は突然変異体 Ty r R S (L W J 1 6) の三次元構造を待つばかりである。

【0275】

突然変異体 Ty r R S (L W J 1 6) を触媒として A T P による O-メチル-L-チロシンとチロシンのアデニル酸形成の動態をピロリン酸交換アッセイにより *in vitro* 分析した。その C 末端に 6 個のヒスチジンを付けた突然変異体 Ty r R S (L W J 1 6) をプラスミド p Q E-60 (Q i a g e n, C A) にクローニングし、プラスミド p Q E-m J Y R S を得た。製造業者のプロトコル (Q i a g e n, C A) に従って固定化金属アフィニティークロマトグラフィーによりタンパク質を精製した。100 mM T r i s H C l (p H 7. 5)、10 mM K F、5 mM M g C l 2、2 mM A T P、2 mM N a P P i、0. 1 mg. mL ウシ血清アルブミン、約 0. 01 μ C i /m L [32 P] N a P P i 及び各種濃度のチロシン又は O-メチル-L-チロシンを含む反応混合物中で 37 °C でピロリン酸 (P P i) 交換を行った。精製突然変異体 Ty r R S (L W J 1 6) を加えて反応を開始し、定期的にアリコートを取分し、0. 2 M N a P P i、7 % 過塩素酸及び 2 % 活性炭で反応停止した。活性炭を濾過し、10 mM N a P P i (p H 2) で洗浄した後、活性炭吸着 A T P 中の 32 P 濃度をシンチレーション計数により測定した。非線形回帰分析を使用してミカエリス・メンテンの式の直接フィットにより *k c a t* 及び *K m* 値を計算した。

【0276】

表 1. リン酸交換アッセイにより測定したチロシンと O-メチル-L-チロシンに対する突然変異体 Ty r R S (L W J 1 6) の動態パラメーター

アミノ酸	<i>K c a t</i> ($10^{-3} s^{-1}$)	<i>K m</i> (μM)	<i>K c a t</i> / <i>K m</i> ($s^{-1} M^{-1}$)
O-メチル-L- L-チロシン	14 ± 1	443 ± 93	32
L-チロシン	1.8 ± 0.2	5833 ± 902	0.31

【0277】

この分析の結果を表 1 に示す。チロシンの *K m* (5833 μM) は O-メチル-L-チロシンの約 13 倍であり、チロシンの *k c a t* ($1.8 \times 10^{-3} s^{-1}$) は O-メチル-L-チロシンの 8 分の 1 である。従って、O-メチル-L-チロシンに対する突然変異体 Ty r R S (L W J 1 6) の *k c a t*/*K m* 値はチロシンの約 100 倍である。大腸菌におけるチロシンの生理的濃度は約 80 μM であり、突然変異体 Ty r R S (L W J 1 6) のチロシンに対する *K m* 値 (5833 μM) よりもはるかに低い。細胞中の O-メチル-L-チロシンの濃度は *K m* (443 μM) と同等以上であると予想される。

【実施例 2】

【0278】

実施例 2 : L-3-(2-ナフチル)アラニンの *in vivo* 組込み

第 2 の非天然アミノ酸 L-3-(2-ナフチル)アラニンを大腸菌でタンパク質への部位特異的組込みを行った。この結果、このスキーム全体を多数のアミノ酸に適用できることが判明した。L-3-(2-ナフチル)アラニンに特異的なシンテターゼは上記実施例 1 で作製した突然変異体 Ty r R S ライブラリーから選択しなかった。

【0279】

非天然アミノ酸をコードするためにアンバー終止コドンとその対応する直交アンバーサプレッサー tRNA である mtRNA_{CUA}^{Tyr} を選択した (Wang, L.; Schultz, P. G. Chem. Biol. 8, 883-890 (2001))。非天然アミノ酸特異性をもつ直交シンテターゼを生産するための出発点として M. jannaschii チロシル-TRNA シンテターゼ (Ty r R S) を使用した。この Ty r R S は内在大腸菌 tRNA をアミノアシル化しない (Steer, B. A.; Schimmel, P. J. Biol. Chem., 274, 35601-35606 (1999)) が、mtRNA_{CUA}^{Tyr} をチロシンでアミノアシル化する (Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. J. Am. Chem. Soc., 122, 5010-5011 (2000))。L-3-(2-ナフチル)アラニンはチロシンによる構造変動が大きく、新規充填特性をもつ可能性があることからこの試験に選択した。

【0280】

mtRNA_{CUA}^{Tyr} に 20 種の共通アミノ酸は負荷せずに L-3-(2-ナフチル)アラニンを負荷するように Ty r R S のアミノ酸特異性を改変するために、M. jannaschii Ty r R S 突然変異体のライブラリーを作製し、スクリーニングした。Bacillus stearothermophilus からの相同 Ty r R S の結晶構造の分析 (Brick, P.; Bhat, T. N.; Blow, D. M. J. Mol. Biol., 208, 83-98 (1989)) に基づき、チロシンのアリール環のパラ位から 7 Å 以内にある M. jannaschii Ty r R S の活性部位の 5 個の残基 (Ty r³²、Asp¹⁵⁸、Ile¹⁵⁹、Leu¹⁶² 及び Ala¹⁶⁷) を突然変異させた。後期選択で野生型シンテターゼ汚染を少なくするために、(Ala¹⁶⁷ を除く) これらの残基をまず全てアラニンに突然変異させた。得られた不活性 Ala₅ Ty r R S 遺伝子を鋳型として使用し、対応する部位にランダム突然変異をもつオリゴヌクレオチドを用いてポリメラーゼ連鎖反応 (PCR) によりランダム突然変異誘発させた。

【0281】

突然変異体 Ty r R S ライブラリーをまずクロラムフェニコールアセチルトランスフェラーゼ (CAT) 遺伝子の非必須位置 (Asp112) のアンバー終止コドンの抑圧に基づくポジティブ選択にかけた。1 mM L-3-(2-ナフチル)アラニンと 80 µg/mL クロラムフェニコールを加えた最少培地で突然変異体 Ty r R S ライブラリーと mtRNA_{CUA}^{Tyr} で形質転換した細胞を増殖させた。細胞は突然変異体 Ty r R S が天然アミノ酸又は L-3-(2-ナフチル)アラニンで mtRNA_{CUA}^{Tyr} をアミノアシル化する場合にしか生存することができない。次に、クロラムフェニコールの存在下で非天然アミノ酸の不在下に生存細胞を増殖させた。生存しなかった細胞は mtRNA_{CUA}^{Tyr} に L-3-(2-ナフチル)アラニンを負荷する突然変異体をコードすると思われるので、非天然アミノ酸を加えたレプリカプレートから抽出した。3 回のポジティブ選択後にネガティブスクリーニングを行った後、CAT 遺伝子での Asp112 TAG コドンの抑圧に基づく *in vivo* アッセイを使用して 4 個の突然変異体 Ty r R S を特性決定した。L-3-(2-ナフチル)アラニンの不在下では、選択した Ty r R S と mtRNA_{CUA}^{Tyr} を発現する細胞は 1% グリセロールと 0.3 mM ロイシンを含む最少培地プレート (GMM L プレート) でクロラムフェニコール濃度 25~35 µg/mL で生存し、L-3-(2-ナフチル)アラニンの存在下では、細胞は GMM L プレートでクロラムフェニコール濃度 100~120 µg/mL で生存した。Ty r R S の不在下 (4 µg/mL クロラムフェニコール) の IC₅₀ 値に比較してこれらの結果は選択した Ty r R S が L-3-(2

ーナフチル) アラニンを受容するが、天然アミノ酸もある程度まで負荷することを示している。

【0282】

天然アミノ酸に対する突然変異体 *TyrRS* の活性を更に弱めるために、上記4種の突然変異体遺伝子を鋳型として使用してDNAシャフリングを1回行った。得られた突然変異体 *TyrRS* ライブラリーを更にポジティブ選択2回とネガティブスクリーニングにかけた。天然アミノ酸に対する活性が大幅に低下し ($IC_{50} = 9 \mu g / mL$ クロラムフェニコール) 且つ *L-3-(2-ナフチル)アラニン* に対する活性が増加 ($IC_{50} = 150 \mu g / mL$ クロラムフェニコール) した1個の突然変異体 *TyrRS (SS12-TyrRS)* が得られた。

10

【0283】

各種突然変異体 *TyrRS* を使用した上記 *in vivo* CATアッセイの結果を表2に示す。*pYC-J17* プラスミドを使用して *mtRNA_{CUA}^{Tyr}* 遺伝子と *Asp112* にアンバー終止コドンをもつクロラムフェニコールアセチルトランスフェラーゼ遺伝子を発現させた。*pBK* プラスミドを使用して *TyrRS* を発現させ、*pYC-J17* と共に大腸菌 *DH10B* にコトランスフェクトした。各種濃度のクロラムフェニコールの存在下に *GMM L* プレートで細胞生存をアッセイした。

【0284】

表2. 突然変異体 *TyrRS* の *in vivo* クロラムフェニコールアセチルトランスフェラーゼアッセイ

20

突然変異体	IC_{50} ($\mu g / mL$ クロラムフェニコール)	
	<i>L-3-(2-ナフチル)-Ala</i> なし	<i>L-3-(2-ナフチル)-Ala</i> 添加
<i>TyrRS</i> なし	4	4
野生型 <i>TyrRS</i>	240	240
選択後		
<i>S1-TyrRS</i>	30	120
<i>S2-TyrRS</i>	30	120
<i>S3-TyrRS</i>	25	110
<i>S4-TyrRS</i>	35	100
DNAシャフリング後		
<i>SS12-TyrRS</i>	9	150

30

40

【0285】

マウスデヒドロ葉酸レダクターゼ (*DHFR*) の *L-3-(2-ナフチル)アラニン* 突然変異体を作製して特性決定し、*mtRNA_{CUA}^{Tyr}* / *SS12-TyrRS* がアンバー終止コドンに応答して *L-3-(2-ナフチル)アラニン* を部位特異的に組込むことができることを確認した。マウス *DHFR* 遺伝子の *Tyr163* コドンを *TAG* に突然変異させ、*DHFR* の *COOH* 末端に *His6* タグを付けて *Ni²⁺* アフィニティークロマトグラ

50

フィーを使用するタンパク質精製を容易にした。陽性対照として野生型M. j a n n a s c h i i Ty r R SをmtRNA_{CUA}^{Tyr}と同時発現させ、TAGコドンでチロシンで効率的に抑圧した(図4)。1mM L-3-(2-ナフチル)アラニンの存在下にSS12-Ty r R SをmtRNA_{CUA}^{Tyr}と同時発現させると、全長マウスDHFRも生産された(液体GMM最少培地中収率2.2mg/L)。L-3-(2-ナフチル)アラニン、mtRNA_{CUA}^{Tyr}又はSS12-Ty r R Sの不在下では、全長DHFRは生産されなかった。ペンターHis抗体を使用してウェスタンブロットでDHFRのCOOH末端のHis6タグを検出した。上記3成分の各々の不在下ではDHFRを検出することができなかった。

【0286】

マウスDHFRのL-3-(2-ナフチル)アラニン突然変異体のトリプシン消化物をMALDI FT-ICRと液体クロマトグラフィータンデム質量分析により分析した処、L-3-(2-ナフチル)アラニンの組込みが明白に確認された。内部校正消化物のペプチドマップはトリプシン消化ペプチドLLPEX* TGVLS EVQEEK(式中、X*はL-3-(2-ナフチル)アラニン残基を表す(Pro164をThrに突然変異させてアンバー抑圧効率を改善した))の理論質量から3.5ppm以内である1867.962に主ピークを示す。更に、所期ペプチドの二重負荷イオンに対応するm/z934.5における前駆体イオンのタンデム質量スペクトルの解釈を図5に示す。スペクトルからの配列情報をまとめると、タンパク質へのL-3-(2-ナフチル)アラニンの部位特異的組込みは明白である。ペプチドマップもLC MS/MS試験もL-3-(2-ナフチル)アラニン残基が他のアミノ酸により置換された突然変異体を示さなかった。ペプチドマップで観察された1500を上回るシグナル対ノイズ比はL-3-(2-ナフチル)アラニンの組込みの忠実度が99.8%を上回ることを示している。

【0287】

得られたSS12-Ty r R Sは突然変異Ty r 32→Leu32、Asp158→Pro158、Ile159→Ala159、Leu162→Gln162、及びAla167→Val167をもつ。B. stea ro t h e r m o p h i l u sからの対応残基はTy r³²(Ty r³⁴)、Asp¹⁵⁸(Asp¹⁷⁶)、Ile¹⁵⁹(Phe¹⁷⁷)、Leu¹⁶²(Leu¹⁸⁰)、及びAla¹⁶⁷(Gln¹⁸⁹)である(括弧内はB. stea ro t h e r m o p h i l u s Ty r R S残基)。

【0288】

相同B. stea ro t h e r m o p h i l u s Ty r R Sの結晶構造によると、Ty r 32→Leu32とAsp158→Pro158の突然変異はTy r 32、Asp158と天然基質チロシン間の水素結合を失わせるのでチロシンとSS12-Ty r R Sの結合に不利であると考えられる。殆どの残基はL-3-(2-ナフチル)アラニンの結合を助長すると予想される疎水性側鎖をもつアミノ酸に突然変異される。

【0289】

要約すると、細胞増殖、タンパク質発現及び質量分析実験の結果、mtRNA_{CUA}^{Tyr}/SS12-Ty r R S対は天然アミノ酸に匹敵する忠実度でアンバーコドンに応答してL-3-(2-ナフチル)アラニンをタンパク質に選択的に挿入できることが判明した。

【実施例3】

【0290】

実施例3：アミノ、イソプロピル又はアリル含有チロシン類似体のin vivo組込み
FACに基づくスクリーニングシステムを使用してアミノ、イソプロピル又はアリル含有チロシン類似体を受容する3種の高度選択性シンテターゼ変異体を迅速に得た。このシステムでは多目的レポータープラスミドを使用してポジティブ及びネガティブ両者の選択圧をかけ、シンテターゼ活性を容易且つ定量的に評価した。クロラムフェニコールアセチルトランスフェラーゼ(CAT)マーカーによりM. j a n n a s c h i iチロシル-tRNAシンテターゼ(Ty r R S)の活性のポジティブ選択を行った。T7ポリメラーゼ/GFPレポーター系により、非天然アミノ酸の存在下と不在下で増殖させた細胞内のシン

10

20

30

40

50

テターゼ活性を評価した。蛍光活性化細胞ソーティング (FACS) を使用して天然アミノ酸を受容するシンテターゼ変異体をスクリーニングし、目視と蛍光分析によりシンテターゼ活性を夫々定性及び定量的に評価した。

【0291】

増幅性蛍光レポーター系の設計。生きた細胞でシンテターゼ活性を評価する汎用スクリーニングシステムを開発する試みは、シンテターゼ変異体集団に適用する選択圧、特にネガティブ選択圧をより精密に制御したいという要望に端を発した。多数のシンテターゼ変異体の活性を評価するためにシステムを使用しなければならないので、高スループットスクリーニングが可能なレポーターが追求された。更に、シンテターゼ活性を容易に定性及び定量評価できるようなレポーターが必要であった。これらの要件を満たすために、蛍光スクリーニングが考案された。このシステムは大腸菌での発現に最適化されたグリーン蛍光タンパク質の変異体である GFPuv のシンテターゼ依存性生産に基づいて考案された (Cramer, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P., *Nature Biotechnol.* 1996, 14, 315-319)。この蛍光体は FACS とフルオロメトリーで使用することができ、更に固体培地と液体培地で目視にも利用できる。このシステムはセクター (例えばアンバーナンセンスコドン) のシンテターゼ依存性抑圧により蛍光シグナルを発生するように設計された。レポーターの感度を最大にするために、他の増幅可能な細胞内レポーター系と同様に GFPuv レポーター遺伝子の発現を誘導するように構成された T7 RNA ポリメラーゼ遺伝子内にアンバーコドン配置することにより増幅可能にした (Lorincz, M., Roederer, M., Diwu, Z., Herzenberg, L. A., Nolan, G. P. *Cytometry*, 1996, 24, 321-329; Zlokarnik, G., Negulescu, P. A., Knapp, T. E., Mere, L., Bures, N., Feng, L., Whitney, M., Roemer, K. & Tsien, R. Y., *Science*, 1998, 279, 84-88)。アンバーコドンを含む T7 RNA ポリメラーゼに合うように RNA 転写産物の生産を容易に最適化できるようにするために、T7 RNA ポリメラーゼ遺伝子をアラビノースプロモーターの制御下においた。

【0292】

T7 RNA ポリメラーゼ / GFPuv レポーター系の最適化。アラビノースプロモーターの制御下にアンバーコドンを含む T7 RNA ポリメラーゼ変異体を発現させると共に T7 プロモーターの制御下に GFPuv 遺伝子を発現させるように中コピーレポータープラスミド pREP を設計した (図 6a)。シンテターゼ依存性蛍光増加を最適化するように設計した 12 種の一連の T7 RNA ポリメラーゼ変異体 (図 6b) を pREP に挿入してプラスミド pREP (1~12) を作製した。全変異体は 7 アミノ酸 (MTMITVH) の N 末端リーダー配列と 1~3 個のアンバー終止コドン (TAG) を含んでいた。変異体 1~3 はリーダー配列のすぐ下流の 1 位の元のメチオニン (M1) を夫々 1、2 及び 3 個のアンバー終止コドンで置換していた。変異体 4~9 は構造のループ領域に位置すると予想される D10、R96、Q107、A159、Q169 又は Q232 に夫々置換するアミノ酸を含んでいた (Jeruzalmi, D. & Steitz, T. A., *EMBO J.*, 1998, 17, 4101-4113)。変異体 10~12 は M1 位で Q107、A159 又は Q232 に夫々置換するアンバー終止コドンを含んでいた。直交グルタミニル tRNA シンテターゼと *S. cerevisiae* 由来グルタミン tRNA_{CUA} を含む適合可能なプラスミドを使用して生きた細胞のフルオロメトリーとフローサイトメトリーによりプラスミド構築物の蛍光増加を検討した。プラスミド pREP (1~12) は各種レベルのシンテターゼ依存性蛍光増加を生じることが判明し、最良構築物 pREP (10) は不活性突然変異体に比較して野生型シンテターゼを含む細胞でフルオロメトリーにより 220 倍の蛍光 (図 6c) を示し、サイトメトリーにより ~400 倍の蛍光中央値 (図 6d) を示した。pREP (10) 内のアンバーコドンに対応する位置で各種官能基を置換した処、T7 RNA ポリメラーゼ内の 107 位の許容性が高いことが判明した

。

【0293】

多目的レポータープラスミドの構築。M. jannaschii TyrRSの変異体を選択及びスクリーニングするために使用する多目的プラスミドを構築するために、プラスミドpREP(10)をプラスミドpYC-J17(Wang, L., Brock, A., Herberich, B. & Schultz, P. G., Science, 2001, 292, 498-500)と融合し、pREP/YC-JUCUAを得た(図7b)。O-メチル-L-チロシン(OMY)の組み込みに選択的なM. jannaschii TyrRSの変異体を発現する適合可能なプラスミド(pBK-mJYRS; Wang, L., Brock, A., Herberich, B. & Schultz, P. G., Science, 2001, 292, 498-500)を使用してプラスミドpREP/YC-JYCUAの機能をアッセイした。pREP/YC-JYCUAとpBK-mJYRSを導入した細胞をOMYの存在下に増殖させた処、プラスミドpYC-J17を使用して得られると同一の120 $\mu\text{g}/\text{ml}$ のクロラムフェニコール(Cm) IC_{50} 値を示し、フルオロメトリーで測定した場合にOMYの存在下に増殖させた細胞の蛍光強度は不在下の330倍であった。

10

【0294】

M. jannaschii チロシル-tRNAシンテターゼの基質特異性の進化。チロシンのフェノール側鎖以外の化学基を選択的に受容するようにM. jannaschii TyrRSのアミノ酸側鎖結合ポケットを進化させることができることが報告されている(Wang, L., Brock, A., Herberich, B. & Schultz, P. G., Science, 2001, 292, 498-500; Wang, L., Brock, A., & Schultz, P. G., J. Am. Chem. Soc. 2002, 124, 1836-1837)。本発明者らはp-イソプロピルフェニルアラニン(pIF)、p-アミノフェニルアラニン(pAF)、p-カルボキシフェニルアラニン(pCF)又はO-アシルチロシン(OAT)の4種の新規官能基(図7b)を酵素に受容させることによりM. jannaschii TyrRSによる非天然アミノ酸組み込みの一般性を更に検討しようと試みた。チロシル環のパラ位に結合ポケットを形成すると考えられる残基であるY32、E107、D158、I159及びL162の位置にランダム変異をもつM. jannaschii TyrRS変異体のライブラリー(Wang, L., Brock, A., Herberich, B. & Schultz, P. G., Science, 2001, 292, 498-500)を、プラスミドpREP/YC-JYCUAを導入した細胞に導入した。ライブラリーダイバーシティ $\sim 10^9$ のこれらの細胞を使用してpIF、pAF、pCF又はOATに選択的なシンテターゼ変異体を同定する4種の進化実験を開始した(図7c)。Cmと4種の非天然アミノ酸の1種の存在下に細胞培養液を飽和まで増殖させることによりポジティブ選択を2サイクル行った。2サイクル目のポジティブ選択後に細胞アリコートを取り出し、アミノ酸又はCmを加えない新しい培養物に接種するために使用し、培養物を再び飽和まで増殖させた。この時点で蛍光発光する細胞は20種の天然アミノ酸の1種を受容できるシンテターゼ変異体を含むと思われる。各株から約 10^8 個の細胞をFACSによるネガティブスクリーニングにかけ、天然アミノ酸を受容するシンテターゼ変異体を排除した。蛍光発光しない細胞を集め、飽和まで増殖させることにより増幅した。これらの増幅細胞を使用して新規培養物に接種し、非天然アミノ酸とCmを加えた液体培地で最終サイクルのポジティブ選択にかけた。飽和まで増殖後に0、30、60又は100 $\mu\text{g}/\text{mL}$ Cmと適当な非天然アミノ酸0又は1 mMを加えた培地に各細胞集団を播種した。

20

30

40

【0295】

進化シンテターゼ変異体の同定と特性決定。pIF、pAF及びOATを加えたCmプレートはアミノ酸を加えないプレートの10~100倍の数の蛍光コロニーを生じた。これに対して、pCF集団のプレートはpCF添加の有無に拘わらず同一数の蛍光コロニーを生じた。非天然アミノ酸を加えたプレートからのpIF、pAF及びOAT集団の各々か

50

ら大きいほうから10個の蛍光コロニーを抽出し、非天然アミノ酸を添加するか又は添加せずに液体培地で飽和まで増殖させた。携帯長波長紫外線ランプを使用して蛍光発生の定性評価を目視により行った(図8a)。

【0296】

蛍光に有意差を生じるクローンに対応するシンテターゼ変異体を配列決定した。pIF及びpAF集団からの全10個のクローンは同一配列であったが、OAT集団からは3種の異なるクローンが同定された。全クローンで5カ所のランダム変異部位にアミノ酸置換があり、pIF-tRNAシンテターゼ(pIF-RS)変異体には更に2箇所の置換があった。各クローンの活性を定量評価した。非天然アミノ酸の存在下又は不在下に液体培地で増殖させた細胞で蛍光をフルオロメトリーにより測定した(図8b)。非天然アミノ酸の存在下又は不在下に各種Cm濃度で細胞をプレーティングすることによりCm IC₅₀を測定した(図8c)。

10

【0297】

4位にアンバーコドンを含むミオグロブリン遺伝子を使用して非天然アミノ酸を含むタンパク質の生産を評価した。pIF、pAF又はOATの存在下又は不在下にpIF-RS、pAF-RS又はOMT-RS変異体を夫々使用して遺伝子を細胞で発現させた(図8d)。タンパク質収率は全3種の変異体で同等であり、非天然アミノ酸含有細胞培養液1リットル当たりタンパク質1~2mgであった。他方、非天然アミノ酸の不在下に増殖させた培養液ではタンパク質生産はほとんど検出できなかった。タンパク質をエレクトロスプレー質量分析により分析した処、夫々の質量はpIF含有タンパク質18457.40±0.81(予想値18457.28)、pAF含有タンパク質18430.30±0.27(予想値18430.21)であった。Cm IC₅₀、フルオロメトリー及びタンパク質発現分析を使用して得られた活性測定値はよく相関していたが、pIF-RSの活性はフルオロメトリーで多少低いようである。他のアッセイに比較して不釣り合いに低いpIF-RA変異体のフルオロメトリー測定値は、レポーター内のアンバー位置の見掛けの許容性にも拘わらずT7 RNAポリメラーゼがpIF類似体の組込み後に部分的に不安定になる可能性を示している。

20

【0298】

多目的レポーター系の有用性。本実施例に記載するレポーター系はポジティブ選択とネガティブスクリーニングの両者に単一の多目的プラスミドを使用できるので、ポジティブ選択とネガティブ選択で交互にプラスミドを交換する必要がなくなる。全体でたった3回のポジティブ選択と1回のネガティブスクリーニングで所望非天然アミノ酸を選択的に受容するシンテターゼ変異体を同定することができた。これらの特徴によりほんの数日間に進化実験を実施することができる。このスクリーニングシステムを使用すると、非天然アミノ酸を加えた寒天プレートを使用して活性シンテターゼ変異体を容易に同定することができ、変異体のアミノ酸特異性を個々にアッセイすることができる。

30

【0299】

上述のように、T7 RNAポリメラーゼ/GFP系はシンテターゼ変異体の活性を定量的に比較するために使用することができる。CATとバルナーゼに基づくポジティブ/ネガティブ選択(表2)を使用して本実施例に記載する3種のOAT-RSクローンと同一ライブラリーから独立して得られる別のOAT-RSクローンを入手できるので、各々に由来するシンテターゼ変異体について2種の異なる進化系を比較することが可能になる(図9)。この分析によると、ポジティブ選択とネガティブスクリーニング後に得られる3種のクローンはOATの存在下ではやや低い蛍光レベルを示すが、非天然アミノ酸の不在下では~10分の1のバックグラウンドレベルを示す。従って、非天然アミノ酸の不在下に対して存在下で増殖させた細胞の蛍光増加は選択とスクリーニング後のOAT-RS(1)を発現する細胞がバルナーゼを使用するポジティブ/ネガティブ選択後のOAT-RSクローンを発現する細胞に比較して約6倍となる。この例が代表例であるか否かははっきりしないが、これらのデータはT7 RNAポリメラーゼ/GFP系が天然アミノ酸基質に対して無差別なシンテターゼ変異体を選択する際にストリンジェンシーを高くできる

40

50

ことを示唆している。しかし、非天然アミノ酸は進化したシンテターゼ変異体との結合を競合して天然アミノ酸の結合が減るので、非天然アミノ酸の不在下よりも存在下で増殖させたほうが細胞の蛍光が増加するのは非天然アミノ酸組込みの忠実度の限界が低くなることに相当すると予想される。更に、忠実度が高いことは明らかに望ましいが、所望用途に応じて忠実度と総シンテターゼ活性のどちらかをとらなければならないと思われる。

【0300】

アミノアシル tRNA シンテターゼ進化の一般性。従来の報告と本明細書に記載する結果から明らかなように、M. jannaschii TyrRS のアミノ酸側鎖結合ポケットは非常に適応性がある。チロシンのフェノール側鎖の代わりに各種官能基を受容するように酵素を進化させることができ、これは高い選択性で実施できる。本願では、チロシン環のパラ位にヒドロキシルの代わりにアミン、イソプロピル又はアリルエーテル官能基を受容するように酵素を進化できることが立証された。

10

【0301】

プラスミド構築。プラスミド pREP (図6a) は T7 RNA ポリメラーゼ遺伝子を含む BamHI / ApaI オーバーラップ PCR フラグメントを rrnB 転写終結領域の上流に挿入した後に pBAD / Myc-HisA プラスミドからの araC 遺伝子及び ara プロモーター領域 (Invitrogen; T7 RNA ポリメラーゼ遺伝子の転写制御用) と GFPuv 遺伝子 (Clontech; T7 ターミネーター領域の上流で T7 プロモーターの下流) を含む ApaI / AhdI オーバーラップ PCR フラグメントをプラスミド pACYC177 (New England Biolabs) の AhaI / BamHI 部位間に挿入することにより構築した。プラスミド pREP (1~12) は位置にアンバー突然変異を含むオーバーラップ PCR フラグメントで T7 RNA ポリメラーゼの HpaI / ApaI フラグメントを置換することにより構築した。プラスミド pREP / YC-JYCUA は pREP (10) からの AfeI / SacII フラグメントと pYC-J17 (Wang, L., Brock, A., Herberich, B. & Schultz, P. G., Science, 2001, 292, 498-500) からの EarI (平滑化) / SacII フラグメントのライゲーションにより構築した。プラスミド pQ を含む細胞に形質転換後に蛍光スクリーニングにより所望構築物を同定した。

20

【0302】

プラスミド pQ は lac プロモーター領域の下流で大腸菌 QRS 終結領域の上流に ScQRS を含む AatII / SalI オーバーラップ PCR フラグメントと、lpp プロモーター領域の下流で rrnC 終結領域の上流に S. cerevisiae tRNA (CUA)⁶¹ⁿ を含む SalI / Aval オーバーラップ PCR フラグメントと、pBR322 (New England Biolabs) の Aval / AatII フラグメントの三重ライゲーションにより構築した。プラスミド pQD は BamHI と BglII の間の pQ フラグメントを ScQRS (D291A) 突然変異体の BamHI / BglII フラグメントで置換することにより構築した。

30

【0303】

プラスミド pBAD / JYAMB-4TAG はプラスミド pYC-J17 (Wang, L., Brock, A., Herberich, B. & Schultz, P. G., Science, 2001, 292, 498-500) と、MjYtRNA_{CUA} の遺伝子と pBAD プロモーターと挿入遺伝子のヘテロ発現用クロニング領域を含む pBAD / Myc-HisA (Invitrogen) のハイブリッドである pBAD / YC-JYCUA プラスミドに C 末端 6 His タグを含むミオグロブリンの S4 アンバー突然変異体の PCR フラグメントを挿入することにより構築した。

40

【0304】

フルオロメトリー及びサイトメトリー分析。所望プラスミドを含む単一コロニーを使用し、適当な抗体と 0.002% アラビノースと必要に応じて適当な非天然アミノ酸を加えた GMM L 培養液 2 mL に接種した。培養液を飽和まで増殖させ、細胞 (200 µL) をペレット化してリン酸緩衝食塩水 (PBS) 1 mL に再懸濁した。600 nm 吸光度により

50

細胞濃度を分析し、FluoroMax-2フルオロメーターを使用して励起波長396 nm、測定波長505 nmで蛍光強度を測定した。PBSに懸濁した細胞をサイトメトリーにより分析した。pREP(10)のT7ポリメラーゼ遺伝子内のアンバー位置の許容性を評価するためには、レポータープラスミドを一連のサブレッサー株に形質転換した後、フルオロメトリーにより分析した。

【0305】

アミノアシル-tRNAシンテターゼ変異体の進化。Y32、E107、D158、I159及びL162の位置をランダム変異させたM. jannaschii TyrRS変異体(Wang, L., Brock, A., Herberich, B. & Schultz, P. G., Science, 2001, 292, 498-500)を、pREP/YC-JYCUAを含むDH10B大腸菌細胞(Life Technologies)に形質転換させ、ダイバーシティ $\sim 10^9$ のライブラリーを作製した。形質転換細胞をSOC培地で60分間37℃で再生させ、LB培地で飽和まで増殖させた。1回目のポジティブ選択を開始するために、ペレット化してGMM L培地に再懸濁したライブラリー培養液2 mLを使用し、25 μ g/mLテトラサイクリン(Tet)、35 μ g/mLカナマイシン(Kn)及び1 mM pIF、pAF、pCF又はOAYを加えたGMM L 500 mLに接種した。37℃で3時間培養後に終濃度75 μ g/mLとなるようにCmを加え、細胞を飽和まで(~ 48 時間)増殖させた。2回目のポジティブ選択では、Tet、Kn、75 μ g/mL Cm及び1 mM pIF、pAF、pCF又はOAYを加えたGMM L 100 mLに1回目のポジティブ選択からの細胞(500 μ L)を接種し、37℃で飽和まで($\sim 24 \sim 36$ 時間)増殖させた。ネガティブスクリーニングに備え、Tet、Kn及び0.02%アラビノース(Ara)を加えたGMM L培地25 mLに2回目のポジティブ選択からの細胞(100 μ L、ペレット化してGMM Lに再懸濁)を接種し、37℃で飽和まで(~ 24 時間)増殖させた。Araで誘導して非天然アミノ酸の不在下に増殖させた細胞(1 mL)をペレット化し、リン酸緩衝食塩水(PBS)3 mLに再懸濁した。Coherent Enterprise IIイオンレーザーを351 nmで励起し、575/25 nm帯域フィルターで発光を検出することによりBDIS FAcVantage TSO細胞ソーターを使用してGFPUVを発現しない細胞を選別した。細胞を集め、TetとKnを加えたLB少なくとも10容量で希釈し、飽和まで増殖させた。3回目のポジティブ選択を開始するために、ネガティブスクリーニングからの細胞100 μ Lをペレット化し、GMM Lに再懸濁したものを使用し、Tet、Kn及び1 mM pIF、pAF、pCF又はOAYを加えたGMM L 25 mLに接種した。3時間37℃で培養後に終濃度75 μ g/mLとなるようにCmを加え、細胞を飽和まで(~ 24 時間)増殖させた。3回目のポジティブ選択後に、Tet、Kn、0.002% Ara、0.75又は100 μ g/mL Cm、及び0又は1 mM pIF、pAF、pCF又はOAYを加えたGMM L/寒天に細胞をプレーティングし、37℃で48時間増殖させた。

【0306】

非天然アミノ酸を含有するタンパク質の発現と特性決定。pBAD/JYAMB-4TAGと適当なpBKプラスミドをコトランスフェクトしたDH10B細胞を使用し、KnとTetを加えて飽和まで増殖させたGMM Lスターター培養液100 mLに接種した。Kn、Tet、0.002% Ara、5 μ M FeCl₃及び所望非天然アミノ酸(又はなし)を加えた培養液500 mLにスターター培養液50 mLを接種し、飽和まで(~ 18 時間)増殖させた。培養液をペレット化し、音波処理し、QiaExpress ionist(Qiagen)Hisタグ精製キットのプロトールに従ってミオグロブリンタンパク質を分離した。タンパク質を12 $\rightarrow$ 20%グラジエントSDSポリアクリルアミドゲル電気泳動とエレクトロスプレー質量分析により分析した。

【実施例4】

【0307】

実施例4：自律21アミノ酸細菌の創製

上述のように、20種の共通アミノ酸は全公知生物で保存されている。しかし、本発明で

は例えば付加機能、構造決定等の拡張遺伝コードを提供する。拡張遺伝コードが例えば特定非天然アミノ酸を組み込んだ細胞に有利であるか否かを調べるためには、所期非天然アミノ酸を生産してこれを組み込む自律細菌が望ましい。本発明はこのような自律21アミノ酸生物を提供し、その成果を付加アミノ酸生物、例えば22アミノ酸生物等の創製に拡張することができる。自律細菌を創製するためには、一般に(i)単純な炭素源から新規アミノ酸を合成できること、(ii)この新規アミノ酸以外は利用しないアミノアシルシンテターゼ、及び(iii)前記シンテターゼ以外ではアシル化されず、他のアミノ酸をコードしないコドンに応答してアミノ酸をタンパク質に送達するtRNAの3点を一般に考慮する。

【0308】

遺伝コードへの新規アミノ酸の*in vivo*組み込みを目的として膨大な研究が進められているが、これらの大半は正常な21アミノ酸細菌を作製するために組み込み特異性をもたない。例えばHest, J. C. M. v., K. L. Kick and D. A. Tirrell, J. Am. Chem. Soc., 2000. 122: p. 1282; Hamano-Takaku, F. ら, J. Biol. Chem., 2000. 275: p. 40324; 及びBudisa, N. ら, FASEB J., 1999. 13: p. 41-51参照。しかし、新規成分を大腸菌の翻訳機構に加え、各種新規アミノ酸を例えば高い忠実度でタンパク質に部位特異的に*in vivo*組み込むことができることが最近示された。例えばWang, L. ら, Science, 2001, 292: p. 498-500; 及びWang, L. とP. G. Schultz, Chem. Comm, 2002: p. 1-10参照。また、2002年4月19日付け同時特許出願(発明の名称"Methods and Compositions for the Production of Orthogonal tRNA-tRNA Synthetase Pairs", 発明者Schultzら)(代理人整理番号54-000130)も参照されたい。

【0309】

本発明は上記技術と生合成経路系を組合せて自律21アミノ酸細菌を創製する。更に、本発明はこのような生物が20種の天然アミノ酸を使用する生物よりも進化的効果があるか又は進化的効果をもつように進化させることができるかという問題に取り組む。

【0310】

完全自律細菌は一般に例えば非天然アミノ酸を生産するための生合成経路系と、非天然アミノ酸を細菌で1種以上のタンパク質に組み込むための翻訳系を含む。翻訳系は一般にこの非天然アミノ酸のみを利用するアミノアシルシンテターゼと、このシンテターゼ以外ではアシル化されず、他のアミノ酸をコードしないコドンに応答して非天然アミノ酸をタンパク質に送達するtRNAを含む。1態様では、改変細菌を最大限に制御できるように生合成経路系遺伝子とアミノアシルシンテターゼ遺伝子とtRNA遺伝子を一般に別々のプラスミドに配置する。

【0311】

1例では、非天然アミノ酸としてp-アミノフェニルアラニン(pAF)を生合成し、タンパク質に*in vivo*組み込む。pAFは場合により例えばその有利な物性(例えばπ供与効果、水素結合特性、弱塩基性、大腸菌に対する毒性の欠如、及び公知二次代謝物であるという事実)に基づいて自律細胞の非天然アミノ酸として選択する。更に、クロラムフェニコールとプリスチナマイシン生産の代謝中間体としてpAFを生産する遺伝子は*Streptomyces Venezuelae*と*Streptomyces pristinaespiralis*で夫々同定されている。例えばYanai, K. ら, *Streptomyces venezuelae* genes *papA*, *papB*, *papC*, PCT国際出願, 2001, 明治製菓株式会社: 日本国. p. 1-83; 及びBlanc, V. ら, *Identification and analysis of genes from Streptomyces pristinaespiralis encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylala*

10

20

30

40

50

nine precursor of pristinamycin 1. Molecular Microbiology, 1997, 23 (2): p. 191-202 参照。上述のように、pAFは場合によりS. Venezuelae酵素PapA、PapB及びPapCと大腸菌アミノトランスフェラーゼを使用して大腸菌でコリスミン酸2（芳香族アミノ酸合成の生合成中間体）から合成する。例えば図15Aに示すようなプラスミドを場合により使用して細胞を形質転換し、それ自体のin vivo pAF供給源を合成する細胞を提供する。pAFのin vivo生合成に使用するプラスミドの1例を配列番号67に示す。配列番号68はコリスミン酸からpAFへの変換を行うために使用される酵素をコードする個々の遺伝子papABCの配列を示す。

【0312】

細胞が非天然アミノ酸（例えばpAF、O-メチル-L-チロシン、グリコシル化アミノ酸、L-dopa等）を生産するように改変されると、この細胞は一般に非天然アミノ酸を細胞内で1種以上のタンパク質に組み込むための翻訳系も付加される。生合成経路系を含むように細胞を改変するのに使用するプラスミドとは別のプラスミドにより翻訳系を細胞に導入すると、例えばコピー数、プロモーター等に関する細胞内のプラスミドの機能をより精密に制御できるので、一般には別のプラスミドを使用する。

【0313】

翻訳機構は一般に例えば2002年4月19日付け同時特許出願（発明の名称"Methods and Compositions for the Production of Orthogonal tRNA-tRNA Synthetase Pairs", 発明者Schultzら）（代理人整理番号54-000130）に記載されているように直交tRNA/R S対を含む。例えば、場合によりMethanococcus jannaschiiチロシル-tRNAシンテターゼ（TyrRS）と突然変異体チロシンアンバーサプレッサーtRNA（mtRNACUATyr）の対を出発点として使用してpAFの直交tRNA/R S対を予備作製する。例えばWang, L. ら, A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. J. Am. Chem. Soc., 2000 122: p. 5010-5011; 及びWang, L. とP. G. Schultz, Chem. and Biol., 2001, 8: 883 参照。

【0314】

例えば、場合によりpAFを受容し且つ20種の共通アミノ酸のいずれをも受容しないようにM. jannaschii TyrRSのアミノ酸特異性を改変することによりpAF特異的シンテターゼ（pAFRS）を作製する。例えばWang, L. ら, Expanding the genetic code of Escherichia coli. Science, 2001, 292: p. 498-500; Wang L. とP. G. Schultz, Expanding the Genetic Code. Chem. Comm., 2002: 1: 1-10; 及びWang, L., A. Brock, and P. G. Schultz, Adding L-3-(2-naphthyl) alanine to the genetic code of E. coli. J. Am. Chem. Soc., 2002. 124: p. 1836 参照。場合によりポジティブ選択とネガティブスクリーニングの組合せを使用し、5カ所（例えばTyr32、Glu107、Asp158、Ile159及びLeu162）にランダムアミノ酸を含むTyr32変異体のライブラリーからpAFrs酵素を同定する。場合により例えば2002年4月19日付け同時特許出願（発明の名称"Methods and Compositions for the Production of Orthogonal tRNA-tRNA Synthetase Pairs", 発明者Schultzら）（代理人整理番号54-000130）に記載されているように、選択とスクリーニングの両者に単一レポータープラスミドを使用する。ポジティブ選択は一般に例えばpAF又は内在アミ

10

20

30

40

50

ノ酸によるクロラムフェニコールアセチルトランスフェラーゼ (CAT) 遺伝子内の許容位置のTAGコドンの抑圧に基づいて行う (例えばWang, L. ら, Expanding the genetic code of Escherichia coli. Science, 2001, 292: p. 498-500; Pasternak, M., T. J. Magliery, and P. G. Schultz, A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code. Helvetica Chimica Acta, 2000 83: p. 2277参照)。pAFを含む液体培地で増殖させたTyrRSライブラリーとレポータープラスミドを含む細胞を一般に例えばクロラムフェニコール (Cm) の存在下での生存に関して選択する。T7 RNAポリメラーゼ遺伝子内の許容位置の2個のUAG終止コドンの抑圧に基づくネガティブスクリーニングはグリーン蛍光タンパク質 (GFP) の発現を誘導する。pAFとCmの不在下で増殖させた細胞をポジティブ選択した後、一般に例えば蛍光活性化細胞ソーティング (FACS) を使用して蛍光の欠如についてスクリーニングする。

10

【0315】

pAFrsの進化: レポータープラスミドpREP(2)/YC-JYCUAはCAT、T7 RNAポリメラーゼ、GFP及びmtRNA_{CUA}^{Tyr}、Tet耐性選択マーカーを含む (Santoro未公開文献)。CAT遺伝子はD112位置にTAGコドン置換を含む。T7 RNAポリメラーゼ遺伝子は7アミノ酸N末端リーダーペプチドとM1及びQ107にTAG置換を含む。ポジティブ選択では、35 µg/ml Kn、25 µg/ml Tet、75 µg/ml Cm及び1 mM pAF (Sigma) を加えたGMM L最少培地で細胞を増殖させた。ネガティブスクリーニングでは、35 µg/ml Kn、25 µg/ml Tet及び0.002%アラビノースを加えたGMM L培地で細胞を増殖させた。Coherent Enterprise IIイオンレーザーでBDIS FACVantage TSO細胞ソーターを使用してFACSを実施した。励起波長は351 nmとし、575/25 nm帯域フィルターで発光を検出した。細胞を集め、TetとKnを加えたLB少なくとも10容量で希釈し、飽和まで増殖させた。

20

【0316】

pAF生合成経路の付加: papA、papB及びpapC遺伝子はS. Venetiae (ATCC10712) ゲノムDNAからPCR増幅した。遺伝子papABCをオーバーラップPCRにより組み立て、pSC101由来プラスミドpLASCに挿入し、アンピシリン選択により維持した。lacZ、malE及びcroの5' UTRからリボソーム結合部位 (rbs) を得、夫々papA、papB及びpapCの前に配置した。papABC遺伝子をlac及びlppプロモーターの制御下におき、2種の経路プラスミドpLASC-lacPW及びpLASC-lppPWを得た。

30

【0317】

pAFRSによるpAF生合成の試験: レポータープラスミド (pREP(2)/YC-JYCUA) とシンテターゼ (pAFRS) と経路プラスミド (pLASC-lacPW及びpLASC-lppPW) の3種のプラスミドを含む大腸菌DH10B細胞をGMM L最少培地で飽和まで増殖させた (バックグラウンド、pAFなし、1 mM外来pAF試験にはpLASCを使用した)。プラスミドを導入せずにDH10Bを増殖させ、レポータープラスミドのバックグラウンド抑圧レベルを測定した。各細胞増殖試料をOD1.0 (600 nm) まで水で希釈し、200 µLをペレット化した。細胞を1% PBS 1 mLに懸濁し、Fluoromax-2 蛍光検出器を使用して分析した (励起波長は351 nmとし、505 nmのピーク発光をモニターした)。DH10Bは1.0 × 10⁴ 蛍光単位を発生したが、レポーター系からのバックグラウンド蛍光 (pAF添加せず) は2.5 × 10⁴ 蛍光単位であった。lacPW、lppPW及び外部から加えた1 mM pAFは夫々7.9 × 10⁴、3.0 × 10⁶、及び3.0 × 10⁴ 蛍光単位を発生した。IPTGはレポータープラスミド (pREP(2)/YC-JYCUA) でアラビノースプロ

40

50

モーターに阻害作用があるため、*lacPW*をIPTGで誘導することはできなかった。

【0318】

芳香族アミノ酸濃度：110 $\mu\text{g}/\text{ml}$ アンピシリンを加えたGMMML最少培地（1%グリセロール、0.3mMロイシン）で*pLASC*プラスミドと*pLASC-lacPW*又は*pLASC-lppPW*を含む大腸菌DH10B細胞を飽和まで増殖させた。外部から*pAF*を加えて増殖させた細胞は増殖開始時に1mMアミノ酸を含んでいた。細胞を遠心分離により回収し（100ml）、洗浄し、水1mlとトルエン0.2mlを加えた。細胞を37℃で30分間震盪した後、遠心分離により分離した。水層を濾過（microcon YM-10）し、HPLC-MS（Agilent 1100）により分析し、1%TFA/アセトニトリル1%TFA（95：5）→（5：95）のグラジエントを用いてZorbax SB-C18カラム（5 μm , 4.6 $\times$ 150mm）で10分間かけて水層5～15 μL を分離した。全イオン質量スペクトルからそのMW（+1）を差し引くことによりアミノ酸を同定した。差し引いたイオンの面積を使用して各試料に存在するアミノ酸の量を計算した。細胞濃度は細胞ペレット中に存在する水分（70重量%）に基づいて計算した。

10

【0319】

*pAF*を含有するタンパク質の発現：テトラサイクリン耐性を付与したプラスミド*pBAD/JYAMB-4TAG*を使用して*lpp*プロモーターと*rrnC*ターミネーターの制御下に*TyrCUAmu*tRNA遺伝子を発現させ、（*Ser4*にアンバー終止コドンをもつ）ミオグロブリン遺伝子をアラビノースプロモーターと*rrnB*ターミネーターの制御下に発現させた。ミオグロブリンのカルボキシ末端には*his6*タグを付けた。*TyrRS*及び*pAFRS*遺伝子はカナマイシン耐性を付与した*pBR322*由来プラスミドで大腸菌*GlnRS*プロモーター及びターミネーターの制御下に発現させた。*papABC*遺伝子は天然ターミネーターの制御下に*pLASC-lacPW*又は*pLASC-lppPW*（13）から発現させた。0.002%アラビノースを加えた最少培地0.5Lでプラスミド*pBAD/JYAMB-4TAG*、*pBK-TyrRS*又は*pBK-pAFRS*と*pLASC*由来プラスミド（指示に応じて*pLASC*、*pLASC-lacPW*又は*pLASC-lppPW*）を含む大腸菌DH10B細胞を増殖させた。外部から*pAF*を加える発現試験は終濃度1mM *pAF*（Sigma）とした。全試験で細胞を飽和まで（20～30時間）37℃で平行して増殖させ、ペレット化し、天然条件下で製造業者のプロトコール（Qiagen, Valencia, Ca）に従ってNi+アフィニティークロマトグラフィーによりタンパク質を精製した。12%SDSポリアクリルアミドゲルで各調製物からの最終タンパク質溶液（3.5ml）から15 μl を分離し、銀染色した。

20

30

【実施例5】

【0320】

実施例5：非天然アミノ酸を生合成するように遺伝子組換えした大腸菌細胞へのO-メチル-L-チロシンのin vivo組込み

本明細書に記載するように、本発明の1側面は大腸菌における非天然アミノ酸の生合成経路である。これは例えば新規酵素の遺伝子を細胞に付加するか又は既存大腸菌経路を改変することにより得られる。本実施例では、非天然アミノ酸O-メチル-L-チロシンを生産するように大腸菌を遺伝子組換えした。

40

【0321】

植物O-メチルトランスフェラーゼはヒドロキシル基をメトキシル基に変換する二次代謝に関与する酵素である。（イソ）オイゲノールO-メチルトランスフェラーゼ（IEMT）とカフェイン酸O-メチルトランスフェラーゼ（COMT）（*Clarkia breweri*）の2種の酵素を大腸菌組込みに選択した。IEMTはオイゲノール/イソオイゲノールをメチル化し、COMTはカフェイン酸をメチル化する。これらの2種の酵素の基質はチロシンに類似する。しかし、これらの酵素はいずれも高い基質特異性とメチル化部位特異性をもつ。

【0322】

50

併用アプローチを使用して両酵素のチロシンに対する基質特異性を進化させ、チロシンをオーメチルーL-チロシンに変換させた。タンパク質の活性部位を突然変異させて大きな突然変異体ライブラリーを作製し、数回選択を行った。3個のクローンが同定された。クローンを特性決定し、少なくとも1個を選択してオーメチルーL-チロシンを生合成する大腸菌株を作製した。上記実施例1に記載したような直交tRNA/R S対も発現するようにこの大腸菌株を遺伝子組換えし、非天然アミノ酸の自律 *in vivo* 組込みを行う細胞を得た。

【実施例6】

【0323】

実施例6：重原子アミノ酸の *in vivo* 組込み

構造に基づく薬剤発見は歴史的に時間と労力のかかるプロセスであり、産業界では薬剤発見プログラムの一端でしか使用されていない。障害の1つはタンパク質構造を決定するためにX線結晶解析を使用する際に遭遇する位相問題である。一般に、セレノメチオニンの存在下でタンパク質を再発現させなければならないので仕事量が倍増する上、必ずしも結晶化に成功するとは限らない。別法は重原子を加えた溶液に結晶を浸漬する方法であるが、結晶が粉碎する恐れがある。重原子含有非天然アミノ酸の *in vivo* タンパク質組込みはタンパク質結晶構造の決定を促進する有用なツールである。

【0324】

p-ヨードフェニルアラニンとp-ブロモフェニルアラニンのタンパク質への部位特異的 *in vivo* 組込みを行った。ヨウ素と臭素は重原子であり、これらを組込みとMADを使用する場合に位相を決定し易くなる。非天然アミノ酸を使用して重原子を部位特異的に導入すると、重原子の位置を選択する際に選択性と柔軟性も得られる。

【0325】

実施例1に記載した方法と組成に従ってp-ヨードフェニルアラニンとp-ブロモフェニルアラニンに夫々特異性をもつ突然変異体シンテターゼを作製した。非天然アミノ酸の *in vivo* 組込みを使用してp-ヨードフェニルアラニンとp-ブロモフェニルアラニンとして臭素又はヨウ素を選択的に導入したタンパク質Zドメイン (B. Nilsson ら, Protein Eng. 1:107-113 (1987)) を発現させた。標準プロトコールに従ってタンパク質結晶トレイをセットした。

【0326】

タンパク質の三次元構造をX線結晶解析により決定し、タンパク質に含まれる重原子を使用して位相を決定した。

【実施例7】

【0327】

実施例7：メタチロシン類似体の *in vivo* 組込み

(実施例1に記載した) mtRNA_{CUA}^{Tyr} をメタチロシン類似体でアミノアシル化するために直交Tyr R Sを作製した。

【0328】

突然変異体Tyr R Sライブラリープラスミドの作製。実施例1に記載した方法にほぼ従って、メタ置換チロシン誘導体に特異的な突然変異体M. jannaschii Tyr R Sをコードするプラスミドのライブラリーを構築した。要約すると、上記実施例1に記載したようにNNKコドンスキームを使用してBacillus stearothermophilus Tyr R Sの結晶構造中の結合チロシンのアリール環のメタ位から6.9 Å以内にあるM. jannaschii Tyr R Sの活性部位の6個の残基 (Tyr³²、Ala⁶⁷、His⁷⁰、Gln¹⁵⁵、Asp¹⁵⁸、Ala¹⁶⁷) をDNAレベルで全20種のアミノ酸に突然変異させた。構築したプラスミドライブラリーpBK-libは約1×10⁹ 個の独立クローンを含んでいた。

【0329】

m-アセチルフェニルアラニン組込みのための直交tRNA-シンテターゼ対の進化。ポジティブ選択3回とネガティブ選択2回を行った後にクロラムフェニコール中での生存が

10

20

30

40

50

非天然アミノ酸付加に依存する5個の候補クローン（配列番号17～21）を得た。m-アセチルフェニルアラニンの不在下では、3種の突然変異体TyrRSプラスミドのうちの1種を含む細胞のクロラムフェニコール耐性の IC_{50} は $20\mu g/ml$ である。m-アセチルフェニルアラニンの存在下では、同一細胞のクロラムフェニコール耐性の IC_{50} は $100\mu g/ml$ である。これらの2つの数値の大差は選択したシンテターゼが細胞内で天然アミノ酸よりもm-アセチルフェニルアラニンを特異的に組込むことができることを示している。m-メトキシフェニルアラニンのデータも同様であり、5個のクローンを分離した（配列番号22～26）。

【0330】

非天然アミノ酸を組込んだDHFRのタンパク質発現。上記で選択したm-メトキシフェニルアラニン及びm-アセチルフェニルアラニンシンテターゼを使用して上記実施例1に記載したようにDHFRでアンバーコドンに応答して該当非天然アミノ酸を組込ませた。陰性対照として、tRNA／シンテターゼの直交対とDHFRをコードするアンバー突然変異体ベクターの両者を含む細胞を非天然アミノ酸の不在下に増殖させた。タンパク質発現の結果を図10に示す。これらの結果は、tRNA／シンテターゼの直交対が非天然m-メトキシフェニルアラニン及びm-アセチルフェニルアラニンを特異的に組込むことを明白に示した。DHFRタンパク質の発現収率はどちらの場合も約 $0.5mg/L$ 培養液であった。

10

【0331】

結合手としてのメタアセチルフェニルアラニンの利用。m-アセチルフェニルアラニンを組込んだDHFRタンパク質を細胞外と細胞内の両方でmg規模のヒドラジド誘導体で標識した。カルボキシル基は水溶液中でヒドラジドと迅速に反応し、生理的条件下で安定なヒドラゾン形成する（Shao, J.; Tam, J. J. Am. Chem. Soc. 117, 3893-3899 (1995)）。この化学を使用してSchultzらはケトンを含む精製T4リゾチームをフルオレセインヒドラジドで特異的に標識している（Cornish, V. W.; Hahn, K. M.; Schultz, P. G. J. Am. Chem. Soc. 118, 8150-8151 (1996)）。

20

【0332】

精製m-アセチルフェニルアラニンを組込んだDHFRタンパク質を水性緩衝液中でフルオレセインヒドラジドで処理した。平行対照として、精製m-メトキシフェニルアラニンを組込んだDHFRタンパク質も同一反応条件にかけた。反応後、両者タンパク質を精製した後、 $491nm$ で励起すると、図11に示す蛍光発光スペクトルが得られた。同一条件下で精製m-アセチルフェニルアラニンを組込んだDHFRはフルオレセインヒドラジドで標識されたが、m-メトキシフェニルアラニンは標識されなかった。

30

【0333】

フルオレセインヒドラジドは細胞浸透性であり、 $4^{\circ}C$ で細胞を溶解しない。従って、m-アセチルフェニルアラニンを組込んだDHFRタンパク質を細胞内でフルオレセインヒドラジドで標識することができる。「ケトン結合手」を組み込んだDHFRを発現する細胞をフルオレセインヒドラジド溶液と共にインキュベートした。 $4^{\circ}C$ で36時間インキュベートし、十分に洗浄して過剰なフルオレセインヒドラジドを除去した後に、標識DHFRタンパク質を精製し、蛍光発光試験にかけた。平行陰性対照として、m-メトキシフェニルアラニンを組込んだDHFRも同一手順で精製した。無傷細胞をフルオレセインヒドラジドで標識した後にDHFRを精製した場合にも細胞外実験（図15）と同様の結果が得られた。

40

【0334】

これらの実験は少なくとも1種の非天然アミノ酸を組込んだタンパク質の有用性の1例を実証するものである。他の化合物を使用して少なくとも1種の非天然アミノ酸を組込んだタンパク質をin vivo標識することもできる。例えばビオチンヒドラジドや他のヒドラジド誘導体が挙げられる。

【実施例8】

50

【0335】

実施例8：光反応性アミノ酸の *in vivo* 組込み

序論：光架橋性アミノ酸を遺伝的にコードし、特定タンパク質に特異的に *in vivo* 組込む実験を行った。その後、光反応基付加用時間制御の励起により必要に応じてこのタンパク質を架橋させた。

【0336】

本発明は例えばタンパク質相互作用の検討に有用である。例えば、本発明はタンパク質中の架橋基の位置を変えることにより各種細胞成分との相互作用を媒介するタンパク質一次配列における残基を決定するのに有用である。タンパク質とこのタンパク質と相互作用する分子の間には共有結合が形成されるので、弱い相互作用や一過的な相互作用を検出することが可能である。

10

【0337】

アリールアジドとベンゾフェノンの2種の化学官能基は300nm（これ以下では光酸化によるタンパク質損傷が問題になることがある）を上回る波長で活性化できるので架橋基として傑出している。これらの2種の架橋基を非天然アミノ酸 *p*-アジドフェニルアラニンと *p*-ベンゾイルフェニルアラニンに夫々組込んだ（図12）。

【0338】

光架橋性アミノ酸に特異的なO-RSの作製。実施例1に記載した直交対である *Methanococcus jannaschii* mtRNA_{CUA}^{Tyr} / TyrRS対を出発点として使用して架橋性非天然アミノ酸 *p*-アジドフェニルアラニン（pBpa）に特異的なO-RSを作製した。突然変異誘発、スクリーニング及び選択方法は実施例1に記載した実験概要に従って実施した。要約すると、5個の残基（Tyr34, Glu107, Asp158, Ile159, Leu162）をランダム突然変異させた突然変異体のMjTyrRSライブラリーを作製した。これらの残基は相同残基（Tyr34, Asn123, Asp176, Phe177, Leu180）が結合チロシンのアリール環のパラ位から6 Å以内にあるアデニル酸チロシルと複合化した *Bacillus stearothermophilus* TyrRSの結晶構造（P. Brick, T. N. Bhat & D. M. Blow *Journal of Molecular Biology* 208, 83（1989））に基づいて選択した。突然変異体TyrRSライブラリーをクロラムフェニコールアセチルトランスフェラーゼ（CAT）遺伝子の許容部位（Asp112）のアンバー終止コドンの抑圧の基づくポジティブ選択にかけた。シンテターゼライブラリーとCAT突然変異体で形質転換した細胞を1mM pBpaとクロラムフェニコールの存在下に増殖させようと試みた。生存細胞は直交mtRNA_{CUA}^{Tyr}に天然又は非天然アミノ酸を負荷することが可能なシンテターゼを含んでいた。mtRNA_{CUA}^{Tyr}と許容部位に3個のアンバー突然変異（Gln2, Asp44, Gly65）を含む毒性バルナーゼタンパク質をコードする遺伝子の変異体（Wang, L., Brock, A., Herberich, B. & Schultz, P. G., *Science*, 292, 498-500（2001））を含む細胞にこれらのシンテターゼ遺伝子を導入した。pBpaの不在下でこれらの細胞を増殖させ、天然アミノ酸を利用することが可能なシンテターゼを選択した。

20

30

40

【0339】

5回のポジティブ及びネガティブ選択後に、全長CAT及びT7 RNAポリメラーゼ（T7 RNAP）の生産が夫々CAT及びT7 RNAP遺伝子中のアンバー終止コドンの抑圧に依存するレポーター株（Santoro SW, Schultz PG. *Proc Natl Acad Sci U S A* Apr 2; 99（7）: 4185-90（2002））に生存シンテターゼプラスミドを形質転換した。T7 RNAPはグリーン蛍光タンパク質（GFP）の発現を誘導するので、これらの細胞をフルオロメトリーによりスクリーニングすることができる。96個のクローンをpBpa依存性クロラムフェニコール耐性とGFP蛍光についてスクリーニングした。6個の別個のシンテターゼが1mM pBpaの存在下と不在下に夫々120mg/Lと5mg/LのIC₅₀で大腸菌

50

に I l e クロラムフェニコール耐性を付与し、同時に p B p a 依存性 G F P 蛍光も示した。p B p a の存在下と不在下のクロラムフェニコール耐性の差は選択したシンテターゼ / t R N A 対がアンバーコドンに应答して細胞に含まれる全 20 種の天然アミノ酸よりも実質的に i n v i v o 特異的に p B p a を挿入することを示している。

【0340】

ミオグロブリンへの p B p a の i n v i v o 組込み。p B p a 組込みの忠実度と効率を測定するために、(C末端 H i s 6 タグを付けた) マッコウクジラミオグロブリンの S e r のコドンをアンバーコドンに変換した。M j p - B p a R S - 1、m t R N A_{C U A}^{T y r} と p B p a の両者の存在下に 2 m g / L の精製収率で全長ミオグロブリンが生産された。p B p a による特異的アンバー抑圧に関与する 3 成分 (アミノ酸、シンテターゼ又は t R N A) のいずれかを除くと、ミオグロブリンの C 末端 H i s 6 タグに対する銀染色又はウェスタンブロットによりミオグロブリンタンパク質を検出することができなかった。このデータは選択したシンテターゼが p B p a に非常に選択的であることを更に裏付けている。

【0341】

突然変異体ミオグロブリンのエレクトロスプレーイオン化イオントラップ質量分析によると p B p a を含むタンパク質の質量計算値である 18519.0 に一致する質量 18159 ± 0.5 が得られた。これは p B p a がタンパク質の単一部位に組込まれたことを裏付けている。天然アミノ酸に対応する質量スペクトルに質量は観察されず、p B p a が高い忠実度で組込まれたことを更に裏付けている。

【0342】

突然変異体 O - R S の配列分析。選択したシンテターゼは興味深い配列収束を示す。M. j a n n a s c h i i T y r R S の T y r 32 は 6 個の突然変異体シンテターゼクローンのうちの 5 個でアラニン又はグリシンに変換される。M. j a n n a s c h i i T y r R S の A s p 158 は選択した 6 個の突然変異体のうちの 5 個でスレオニンに変換され、I l e 159 は 6 個の突然変異体のうちの 4 個でセリンに変換される。セリン又はプロリン置換は M. j a n n a s c h i i T y r R S の 107 位に位置し、L e u 162 は 6 個の突然変異体のうちの 4 個で保存されている。この分析から突然変異のコンセンサス組合せ (32 : G l y, A l a / 107 : S e r, P r o / 158 : T h r / 159 : S e r / 162 : L e u) が判明した。

【0343】

G S T への p B p a の i n v i v o 組込み。タンパク質-タンパク質相互作用のマッピングにおけるこの方法の有用性を立証するために、グルタチオン-S-トランスフェラーゼを用いて架橋実験を実施した。このタンパク質はグルタルアルデヒドを使用して予め非特異的に架橋されている 2 個の同一サブユニットのダイマーである。ダイマー S c h i s t o s o m a J a p o n i c a グルタチオン-S-トランスフェラーゼ (S j G S T) の結晶構造 (M c T i g u e, M. A., W i l l i a m s, D. R. & T a i n e r, J. A. Journal of Molecular Biology 246, 21-27 (1995)) を使用して p B p a で置換する 2 部位、即ち結晶構造のダイマー界面に隠れている残基 P h e 52 と溶媒に暴露されている残基 T y r 198 を同定した。27 k D a タンパク質 S j G S T の遺伝子における P h e 52 又は T y r 198 に対応するコドンをアンバーコドンで置換した。次に、直交シンテターゼ t R N A 対を使用し、大腸菌で S j G S T のこれらの部位に p B p a を部位特異的に組込んだ。長波長紫外線を照射すると、精製 S j G S T は共有結合ホモダイマーに変換されることが変性 S D S P A G E により分かった。存在する S j G S T の約 70% が 5 分間で架橋した。これに対して、野生型 S j G S T 又はダイマー界面の外側に位置する残基 198 に p B p a を組込んだ S j G S T を使用した対照実験では UV 照射に应答する架橋を検出できない。

【0344】

これらの結果は、タンパク質-タンパク質相互作用に関与するアミノ酸を決定するために部位特異的 p B p a 置換を使用できることを示している。

【0345】

突然変異体シンテターゼの特性決定。DH10B／pREP(2)／YC-JYCUAにおける各シンテターゼクローンを使用し、カナマイシンとテトラサイクリンを30、20 mg／Lまで加えたLB0.5 mLに接種した。20時間増殖(37℃、300 rpm)後に細胞をdH₂Oで10⁴倍に希釈し、レプリカを2組のGMMMLプレートにスポットした。第1組のプレートにはカナマイシン及びテトラサイクリン夫々30及び20 µg／Lと濃度0 µg／L～110 µg／Lのクロラムフェニコールを加えた。第2組のプレートは1 mM pBpaを加えた以外は第1組と同一とした。48時間後に、クロラムフェニコールを加えないプレート上のコロニーの半数が見えるクロラムフェニコール濃度からpBpaの存在下と不在下のクロラムフェニコール耐性のIC₅₀を計算した。Storm 10
ホスホイメジャー(Molecular dynamics)を使用してpBpaの存在下と不在下のGFP発現を画像化した。GFPシグナルとクロラムフェニコール耐性の両者に最大のアミノ酸依存性を示す突然変異体シンテターゼ遺伝子を標準方法により分離し、配列決定した。

【0346】

タンパク質発現。アラビノースプロモーターとrrnBターミネーターにつなげた突然変異体SjGST遺伝子と、lppプロモーターとrrnCターミネーターにつなげたmtRNA_{CUA}^{Tyr}と、テトラサイクリン耐性マーカーを含むプラスミドPYC／SjGSTmutを、p-BpaRSを発現するpBKベクターと共にDH10B大腸菌にコトランスフェクトした。カナマイシン30 µg／Lとテトラサイクリン25 µg／Lを加えた2 x-YT10 mL中で細胞を増幅した後、PBSで洗浄したものを使用し、適当な抗生物質とpBpa1 mMまでを加えた液体GMMML1 Lに接種した。アラビノース0.2%までを加えてから5時間増殖させることによりOD₆₀₀ = 0.6でタンパク質発現を誘導した。細胞を遠心分離により回収し、Ni-NTAアフィニティークロマトグラフィーを使用してC末端ヘキサヒスチジンタグによりタンパク質を精製した。 20

【0347】

0.002%アラビノースで誘導した以外はSjGSTと同様の方法でpBAD／JYAMB-4TAGを含む細胞からマッコウクジラミオグロブリンを発現させ、精製した。質量分析用試料をNAP-10カラム(Pharmacia)で脱塩し、HPLCにより精製した。pBpaの組込みを確認するために、エレクトロスプレーイオン化イオントラップ質量分析によりタンパク質量を確認した。 30

【0348】

突然変異体SjGSTクローニング。pGEX-3(Pharmacia)を鋳型としてオーバーラップPCRにより突然変異体SjGST遺伝子を組み立てた。全PCR反応は製造業者の指示に従ってExpand PCRキット(Roche)を使用して実施した。得られた遺伝子をNcoI及びKpnI制限酵素で消化し、予め消化して脱リン酸しておいたpBADJYCベクターの同一制限部位間にC末端ヘキサヒスチジンタグとインフレームでクローニングした。全最終構築物はDNAシーケンシングにより確認した。

【0349】

光架橋。架橋反応は(50 mM NaH₂PO₄, 300 mM NaCl, 250 mM イミダゾール中)10 ng／µL SjGST100 µLを使用して96穴マイクロタイタープレート(Nuncsorb)で4℃で実施した。携帯UVランプ(115 V, 60 Hz, 0.2 A; Spectronics, NY, USA)を使用して365 nmで試料に1分間又は5分間照射した。試料をウェルから取出し、SDSローディング緩衝液で希釈し、10→20%グラジエントゲルでSDS-PAGEにより生成物を分解した。SjGSTをPVDF(Biorad)に移し、ヤギ抗GST(Pharmacia)と二次マウス抗ヤギHRPコンジュゲート(Sigma)を使用してウェスタンブロットにより探索した。Super Signal West(Pierce)を使用してシグナルを発生させ、ハイパーフィルム(Amersham)で露光して可視化した。 40

【実施例9】

【0350】

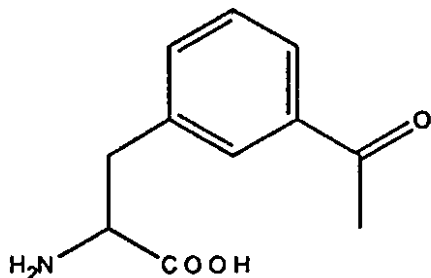
実施例9：メタ置換フェニルアラニンの合成

1 側面では、本発明は式 I V：

【0351】

【化4】

IV



10

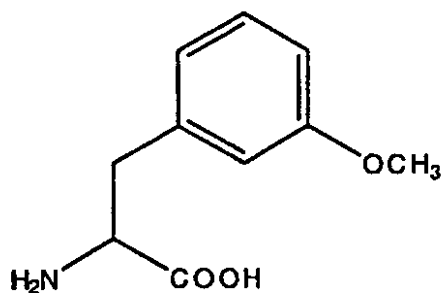
【0352】

及び式 V：

【0353】

【化5】

V



30

に示すメタ置換フェニルアラニンを提供する。

【0354】

式 I V は 3-アセチルフェニルアラニンの構造を示し、式 V は 3-メトキシフェニルアラニンを示す。

40

【0355】

メタ置換フェニルアラニンは図 1 4 に要約するような手順で合成される。一般に、NBS (N-ブロモスクシンイミド) をメタ置換メチルベンゼン化合物に加えてメタ置換臭化ベンジルを得た後、マロン酸化合物と反応させてメタ置換フェニルアラニンを得る。メタ位に使用する典型的置換基としてはケトン、メトキシ基、アルキル、アセチル等が挙げられるが、これらに限定されない。特定例を以下に示す。

【0356】

NBS (N-ブロモスクシンイミド) を使用前に沸騰水から再結晶させた。NBS (1.85 g, 10.5 mmol) を 3-メチルアセトフェノン (1.34 g, 10 mmol) を

50

の溶液に加えた。AIBN（2', 2'-アゾビスイソブチロニトリル）（0.043 g, 0.25 mmol）を混合物に加えた。反応混合物を4時間還流させた。反応の完了をTLC（8:1／ヘキサン：EtOAc）により確認した。水を加えて反応を停止した後、有機溶媒を除去し、ヘキサンを加えて固形分を得た。固形分を濾過し、ヘキサンとEtOAcで洗浄した。次に、混合物をヘキサンで再結晶させた。上清を集め、溶媒を除去し、化合物（1-（3-ブロモメチルフェニル）エタノン）を得た。

【0357】

ペンタンで洗浄したナトリウム片（2.3 g, 0.1 mol）にアルゴン雰囲気下で無水エタノール（50 ml）を滴下した。滴下の完了後に、ナトリウム片を完全に溶解させるには攪拌が必要であった。ジエチルアセチルアミドマロン酸エステル（21.7 g, 0.1 mol）の溶液を30分間かけて加えた。無水エタノール中の1-（3-ブロモエチルフェニル）エタノン（21.1 g, 0.1 mol）を90分間かけて滴下した。混合物を一晩還流後にエーテルと水を加え、有機層を分離した。水を加えて反応を停止した後、有機層を合わせてブラインで洗浄し、MgSO₄で乾燥し、濾過した。溶媒を減圧除去した。ヘキサン-ジクロロメタン4:1を残渣に加え、不溶分を濾去し、10:1ジクロロメタン-ベンゼンで完全に洗浄し、2-アセトアミド-2-[3-アセチルフェニル]メチルマロン酸ジエチルを得た。この化合物をジオキサン中で8M HClと共に一晩攪拌した。次に、混合物を乾燥し、水を加え、再び乾燥し、最終化合物塩酸m-アセチルフェニルアラニンを得た。HPLCを使用して所望化合物を精製し、白色固体とした。総収率は64%であった。¹H NMR（D₂O）：δ 7.85-7.28（m, 4H），4.23（dd, 1H），3.2（m, 2H），2.7（s, 3H）。分子量計算値243.69、実測値243.07。同様の合成を使用して3-メトキシフェニルアラニンを製造した。この場合の臭化ベンジルのメタ位のR基は-OCH₃である。例えばMatsoukasら, J. Med. Chem., 1995, 38, 4660-4669参照。

【実施例10】

【0358】

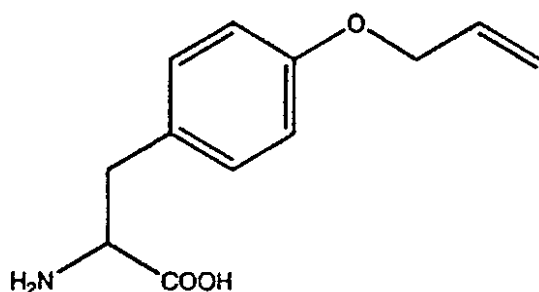
実施例10：4-アリル-L-チロシンの合成

別の側面では、本発明は式II：

【0359】

【化6】

II



に示す構造の4-アリル-L-チロシンを提供する。

【0360】

式IIの化合物4-アリル-L-チロシンは図13に示すスキームに従って合成される。保護チロシン（例えばNboC又はFmoc保護チロシン）を臭化アリルと反応させて保護アリルチロシンを得た後、一般に脱保護して4-アリル-L-チロシンを得る。1例で

は、N-(tert-ブトキシカルボニル)-L-チロシン (2.95 g, 10 mmol) を DMF 80 ml に溶かした。溶液を 5℃ まで冷却し、NaH (0.63 g, 26 mmol) を加えた。反応混合物を 10℃ まで昇温させ、更に 2 時間攪拌した。その後、臭化アリル (1.33 g, 11 mmol) を混合物に加え、反応混合物を 10℃ まで昇温させ、更に 2 時間攪拌した。その後、混合物に臭化アリル (1.33 g, 11 mmol) を加え、反応混合物を室温まで昇温させた。反応混合物を 4 時間攪拌した。水を加えて反応を停止した。水層を酢酸エチルと CH₂Cl₂ で抽出した。有機層を無水 MgSO₄ で乾燥した。有機溶媒を除去し、白色固体を得た。次にこの化合物を 1,4-ジオキサン中 4 M HCl 中で 4 時間還流した。全溶媒を蒸発させ、所望生成物を白色固体として得た (1.9 g, 86%)。¹H NMR (CD₃OD): δ ppm 3.1 (m, 2H), 4.1 (t, 1H), 4.5 (d, 2H), 5.3 (q, 1H), 5.9 (m, 1H), 6.9 (d, 2H), 7.1 (d, 2H)。分子量計算値 221、実測値 222。

10

【実施例 11】

【0361】

実施例 11：非天然アミノ酸の細胞取込みスクリーニング

市販品又は市販品を出発材料として簡単に合成した各種非天然アミノ酸と所期 α-ヒドロキシ酸 (I. Shin, B. Herberich, A. Varak, T. Maglier, P. Schultz, 未公開文献) を細胞毒性についてスクリーニングした。例えば、図 29 は以下のスクリーニングに有用な非天然アミノ酸のライブラリーを示す。各アミノ酸をグリセロール最少培地中 1 mM で (例えば pBLAM-YQRS と pACYsupA38 を含む DH10B に対する) 細胞毒性についてスクリーニングした。毒性は (1) 倍加時間の有意変化がない無毒性、(2) 倍加時間の増加が約 10% 未満である低毒性 (図 29 の化合物 S63、S69、S74、S75、S81、S95)、(3) 倍加時間の増加が約 10% ~ 約 50% である中毒性 (図 29 の化合物 B、M、P、S12、S14、S22、S41、S45、S49、S52、S62、S64、S65、S71、S91、S93、B10)、(4) 倍加時間の増加が約 50% ~ 約 100% である高毒性 (図 29 の化合物 C、Q、V、BB、S2、S5、S50、S60、S78、S83、S89、S90)、及び (5) 倍加時間の増加が約 100% を上回る極毒性 (図 29 の化合物 W、S15、S26、S27、S30、S31、S39、S47、S88、S94) の 5 分類に分類される。例えば Liu, D. R. & Schultz, P. G. Progress toward the evolution of an organism with an expanded genetic code. Proceedings of the National Academy of Sciences of the United States of America 96, 4780-4785 (1999) 参照。

20

30

【0362】

高毒性又は極毒性に分類されるアミノ酸の毒性は、一般に IC₅₀ 値を得るためにその濃度の関数として測定される。一般に、天然アミノ酸の近縁類似体 (例えば Q、W、S5、S26、S27、S50、S90、S94) であるか又は反応性官能性を示す (例えば S15、S39、S47) アミノ酸が最高毒性を示す。

40

【0363】

毒性アミノ酸の可能な取込み経路を同定するために、例えば過剰 (2 mM) の構造的に類似する天然アミノ酸を加えた培地で IC₅₀ レベル (一般に 3 μM ~ 500 μM) で毒性アッセイを繰返した。毒性アミノ酸では過剰の天然アミノ酸の存在は毒素存在下の細胞の増殖能を救済したが、これは天然アミノ酸が細胞取込み又は必須酵素との結合のために毒素と有効に競合解消したためであると予想される。これらの場合には、毒性アミノ酸を可能な取込み経路に割り当て、細胞生存にその相補が必要な「致死対立遺伝子」と呼ぶことができる。こうして同定された致死対立遺伝子 (毒性非天然アミノ酸の 16 種) はアラニン、グルタミン酸、リジン、ロイシン、メチオニン、プロリン、グルタミン、アルギニン

50

、スレオニン及びチロシンの10種の可能なアミノ酸取込み基に及ぶ。

【0364】

これらの致死対立遺伝子は細胞が非毒性非天然アミノ酸を取込む能力をアッセイするのに極めて有用である。IC50レベルの各対立遺伝子を加えた培地に各非毒性非天然アミノ酸2mMを加えた。毒性対立遺伝子の相補は細胞増殖の回復により判断され、恐らく致死対立遺伝子に割り当てられると同一の取込み経路により非毒性アミノ酸が細胞に取込まれることを示す。相補の欠如については結論が出ていない。

【0365】

この方法を使用し、22種のグルタミン及びグルタミン酸類似体がDH10Bに取込まれるか否かを調べた。アミノ酸S27及びS47を毒性グルタミン対立遺伝子として夫々100μMと30μMで使用し、S50を毒性グルタミン酸対立遺伝子として150μMを使用した。S27とS47の相補の結果は完全に一致し、(S27とS47に加えて)アミノ酸B、Z、S6、S60、S61及びS62も恐らくグルタミン取込み経路を介して細胞に取込まれることが判明した。同様に、S50の相補からB、C、K、X、S60、S65及びS84が恐らくグルタミン酸輸送系を介してDH10Bに取込まれることが判明した。

【0366】

これらの知見は大腸菌グルタミン及びグルタミン酸輸送経路が側鎖延長(X及びZ)、γ位にケトン又はメチレンを配置(B、C、S65)、カルボキサミドをスルホキシド(S61)、細菌グルタミン輸送体又はヒドラジドの公知基質(S47)、更にグルタミン輸送体基質で置換及び側鎖末端の各種ハイブリダイゼーション置換(S60、S62、K、S84)等のアミノ酸構造の有意変動に耐えられることを示している。Jucovic, M. & Hartley, R. W. Protein-protein interaction: a genetic selection for compensating mutations at the barnase-barstar interface. *Proceedings of the National Academy of Sciences of the United States of America* 93, 2343-2347 (1996) 及び Weiner, J. H., Furlong, C. E. & Heppel, L. A. A binding protein for L-glutamine and its relation to active transport in *E. coli*. *Archives of Biochemistry and Biophysics* 142, 715-7 (1971) 参照。

【実施例12】

【0367】

実施例12：p-アミノフェニルアラニンの生合成

非天然アミノ酸p-アミノフェニルアラニン(pAF)をin vivo生産させるためには、場合によりクロラムフェニコールとプリスチナマイシン生産経路に依存する遺伝子を使用する。例えば、*Streptomyces Venezuelae*と*Streptomyces pristinaespiralis*でこれらの遺伝子は代謝中間体としてpAFを生産する。例えばYanai, K. ら, *Streptomyces venezuelae* genes papA, papB, papC, PCT国際出願, 2001, 明治製菓株式会社：日本国. p. 1-83; 及びBlanc, V. ら, *Identification and analysis of genes from Streptomyces pristinaespiralis encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinomycin* 1. *Molecular Microbiology*, 1997, 23 (2): p. 191-202 参照。

【0368】

pAFの生合成経路を図15Bに示す。pAFは場合により芳香族アミノ酸合成の生合成

中間体であるコリスミン酸（図15Bの化合物2）から大腸菌で合成する。コリスミン酸からpAFを合成するためには、細胞は一般にコリスミンシンターゼ、コリスミン酸ムターゼ、デヒドロゲナーゼ（例えばプレフェン酸デヒドロゲナーゼ）及びアミノトランスフェラーゼを使用する。例えば、（例えば図15Bに示すように）*S. Venezuela* 酵素PapA、PapB及びPapCと大腸菌アミノトランスフェラーゼを使用し、コリスミン酸からpAFを生産する。

【0369】

例えば、4-アミノ-4-デオキシコリスミン酸シンターゼは例えば単純な付加／脱離反応で（グルタミンからの）アンモニアを使用してコリスミン酸を4-アミノ-4-デオキシコリスミン酸（図15Bの化合物3）に変換する。コリスミン酸ムターゼとプレフェン酸デヒドロゲナーゼに夫々類似するPapBとPapCを使用し、4-アミノ-4-デオキシコリスミン酸を4-アミノ-4-デオキシプレフェン酸（図15Bの化合物4）に変換した後、p-アミノフェニルピルビン酸（図15Bの化合物5）に変換する。例えば大腸菌に由来する非特異的チロシンアミノトランスフェラーゼを使用してp-アミノフェニルピルビン酸をpAFに変換する。例えば*Escherichia coli* and *Salmonella*, 第2版, F. C. Neidhardt編, Vol. 1. 1996, Washington, D. D. : ASM Press参照。例えば、場合によりtyrB、aspS又はilvEを使用してp-アミノフェニルピルビン酸からpAFを生産する。

10

【0370】

図13はpAFの生合成で使用するプラスミドを示す。図のプラスミドは例えばlac又はlppプロモーターの制御下にpSC101由来pLASCプラスミドにクローニングされた*S. venezuela* 遺伝子papA、papB及びpapCを含む。このプラスミドを使用し、遺伝子によりコードされる酵素を生産するように細胞（例えば細菌細胞）を形質転換する。発現されると、酵素は所望非天然アミノ酸（例えばpAF）の生産を目的とする1種以上の反応を触媒する。例えば、タンパク質PapA、PapB及びPapCはコリスミン酸をp-アミノフェニルピルビン酸に変換し、大腸菌芳香族アミノトランスフェラーゼは生合成を完了し、pAFを生じる。

20

【0371】

一般に、本発明におけるコリスミン酸からのpAFの合成は細胞で生産される他のアミノ酸（例えばコリスミン酸から一般に生産される他の芳香族アミノ酸）の濃度に影響を与えない。一般に、p-アミノフェニルアラニンは効率的タンパク質生合成に十分な濃度（例えば天然細胞量）で生産されるが、他の芳香族アミノ酸の濃度を変化させたり細胞資源を枯渇させる程にはならない。こうしてin vivo生産されるpAFの典型濃度は約10mM～約0.05mMである。*S. venezuela*での知見は、シキミン酸経路の調節が第4の芳香族アミノ酸の生産におけるコリスミン酸消費を補償するように改変されることを示唆している。例えばHe, J.ら, *Microbiology*, 2001 147: p. 2817-2829参照。上記経路で使用される酵素を生産するために使用される遺伝子を含むプラスミドで細菌を形質転換し、21番目のアミノ酸としてpAFを作製したら、場合によりin vivo選択を使用してリボソームタンパク質合成と細胞増殖の両者に合うようにpAFの生産を更に最適化する。

30

40

【0372】

pAF tRNA-シンテターゼ対はタンパク質の非必須位置におけるTAGコードンの抑圧を許容するので、場合によりこのタンパク質の生産により生合成経路の有効性をモニターし、最適化する。タンパク質生合成に十分な濃度のpAFを生産する細胞のみがTAGコードンを抑圧することが可能である。同時に、TAG-タンパク質が細胞増殖に必須のタンパク質である場合には大腸菌増殖率に基づいて最適pAF生産を選択することができる。生合成遺伝子をプラスミドに配置すると、例えばプラスミドコピー数とプロモーター強度を変えることにより、pAF生産レベルを改変することができる。pAF生合成経路の付加が大腸菌における他の芳香族アミノ酸の生産に影響を与えるか否かを調べ、pAF生

50

産を定量するためには、場合により例えば抽出とLCMS分析により芳香族アミノ酸の細胞濃度をモニターする。例えばMoss, R. E., Methods in Enzymology, 1995. 262: p. 497-499及びMimura, H., S. Nagata, and T. Matsumoto, Biosci. Biotech. Biochem., 1994, 58(10): p. 1873-1874参照。

【実施例13】

【0373】

実施例13: Dopaの生合成

dopaをin vivo生合成するためには、例えば大腸菌に由来する非特異的芳香族ヒドロキシラーゼの1種以上の遺伝子(例えばhpaBC)を一般にlppプロモーターの制御下においた低コピー数ベクター(例えばpSC101誘導体)にクローニングする。この構築物は図20に示すように、チロシン(1)からdopa(2)をin vivo生産するが、増殖細胞に非毒性である。この遺伝子を用いて同様に操作し、精製目的でdopaを過剰生産させた。例えばJang-Young Lee, Luying Xu, Biotechnology Letters, 1998, 20, 479-482参照。しかし、上述のように、過剰生産は一般に望ましくない。本願では、低コピープラスミドを使用して天然細胞量のdopaを生産させる。

10

【実施例14】

【0374】

実施例14: O-メチル-L-チロシンの生合成

O-メチル-L-チロシンは場合によりヒドロキシル基をメトキシル基に変換する二次代謝に関与する酵素である植物O-メチルトランスフェラーゼにより生合成する。このような酵素として(イソ)オイゲノールO-メチルトランスフェラーゼ(IEMT)とカフェイン酸O-メチルトランスフェラーゼ(COMT)の2種を選択した。いずれもClarkia breweriに由来する。IEMTはオイゲノール/イソオイゲノールをメチル化し、COMTはカフェイン酸をメチル化する。これらの2種の酵素の基質はチロシンに類似する。しかし、これらの酵素はいずれも高い基質特異性とメチル化部位特異性をもつ。従って、これらの2種の酵素がチロシンをその基質として使用してチロシンをO-メチル-L-チロシンに変換するようにこれらの酵素を進化させる併用アプローチを採用した。活性部位残基を突然変異に選択し、大きな突然変異体ライブラリーを作製した。数回選択後に少なくとも約3個のクローンが同定された。

20

30

【0375】

他の態様では、O-メチル-L-チロシンを生産するために使用する酵素を人工的に進化させることもでき、例えば式IIIに示すようなメタ置換メトキシフェニルアラニンを生産するように進化させることもできる。

【実施例15】

【0376】

実施例15: グリコシル化アミノ酸の生合成

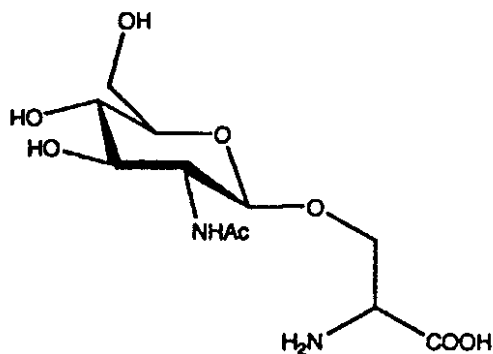
本発明はグリコシル化アミノ酸の生合成法も提供する。グリコシル化アミノ酸のin vivo形成は場合により多数の方法で実施される。例えばN-アセチルガラクトサミニダーゼ、トランスグリコシラーゼ、又はヒドロラーゼ(例えば逆方向に作用する例えばセリナーゼ)を含むプラスミドで細胞を形質転換すると、グリコシル化アミノ酸を生産する細胞が得られる。下記に記載するような翻訳系と組合せると、この生合成経路の結果、グリコシル化アミノ酸を生産して細胞内で1種以上のタンパク質に組込む細胞が得られる。例えば、図28はグリコシル化アミノ酸の形成を示しており、Rは場合によりアルコール、アミン、又はN-アセチルアミンである。構造の1例を式IV:

40

【0377】

【化7】

IV



10

に示す。

【実施例 16】

【0378】

実施例 16：非天然アミノ酸の組込みによる効果の識別

20

例えば DNA 内のナンセンスコドンに応答して高い翻訳効率と忠実度で基本炭素源から非天然アミノ酸を生合成し、このアミノ酸をタンパク質に組込むことが可能な完全自律細菌を本発明により開発できるとしても、この細菌がこのような付加により実際に 20 種の天然アミノ酸のみを組込む生物にまさる効果を得るのかどうかは問題である。本発明は拡張遺伝コードがこのような効果を提供するか否かを調べると共に、効果の種類と効果の元となる非天然アミノ酸を識別する方法を提供する。

【0379】

19 世紀以来、細菌学者は増殖条件を変えると細菌培養物が大きく変化することに関心を示してきた。例えば Summers, W. C., J. Hist. Biol., 1991, 24: P. 171-190 参照。しかし、20 アミノ酸生物では全進化形態が検討されて

30

いる。本発明は非天然アミノ酸を組込んだ大腸菌の遺伝コードの拡張可能性について検討し、付加アミノ酸を組込む能力が大腸菌に進化的効果を与えるか否を試験する方法を提供する。

【0380】

遺伝コードに新規アミノ酸を付加することにより大腸菌に進化的効果が得られるか否かを調べるためには、場合により 21 アミノ酸細菌の進化を 20 アミノ酸細菌の進化と比較する。このアプローチは非天然アミノ酸をタンパク質に組込む新規翻訳機構群と選択圧下においた突然変異誘発大腸菌ゲノムライブラリーを組合せるものである。上記及び Schulltzらが他の文献に記載している遺伝的選択アプローチによると、TAG コドンに

40

応答して高い忠実度で効率的に新規アミノ酸をタンパク質に組込むことができる少なくとも約 11 種の新規アミノアシルシンテターゼがこれまでに作製されている。

【0381】

大腸菌ゲノムのプラスミドライブラリーを突然変異誘発すると、コドンが入り乱れ、ゲノムに TAG ナンセンスコドンがランダムに付加される。新規 TAG コドンは発現されるタンパク質に新規アミノ酸を組込むことにより抑圧することができる。これらの細菌系に選択圧をかけ、大腸菌増殖亢進について選択する。例えば図 21 参照。効果を付与するライブラリーから選択したゲノムフラグメントを場合により分離し、TAG コドンに

応答して他の非天然アミノ酸とチロシンを組込む場合には増殖能の亢進についてスクリーニングする。

【0382】

50

場合により pSC101 低コピーベクター（約 5 コピー／細胞）を使用して当業者に公知の標準方法により大きなインサート（7～14 kb）大腸菌ゲノムライブラリーを構築する。例えば、個体数 600 の pSC101 由来大腸菌ゲノムライブラリーは大腸菌ゲノムを完全に網羅すると共にアミノアシルシンテターゼと上記 tRNA プラスミドに適合可能である。多数の突然変異原が TAG コドンで遺伝子に組込む能力について検討されている。例えば Miller, J. H., *A short course in bacterial genetics*, 1992, Plainview: Cold Spring Harbor Laboratory Press 参照。各突然変異原はその TAG コドン形成に関して完全にランダムではないので、場合により多重突然変異誘発法を使用する。同一の個体数 600 の大腸菌ゲノムライブラリーを 4 種の異なる突然変異原で突然変異誘発することにより、TAG コドンを一般できるだけ多数の部位に配置する。ゲノムライブラリーを突然変異させ、ライブラリーを合わせて個体数 > 1010 の 1 つの大きな突然変異ゲノムライブラリーを作製するためには、紫外線照射、ミューテーター株 (XL1 red)、4-ニトロキノリン-1-オキシド (NQO) 及びエチル・メタン・スルホン酸 (EMS) の 4 種の方法を選択できるが、これらに限定されない。これらの突然変異法はいずれも主に点突然変異の形成に依存するが、突然変異誘発メカニズムで補完し合う結果、全体ではより均一な TAG コドン分布が得られる。紫外線照射とミューテーター株は全塩基置換を誘発するが、NQO と EMS は基本的に G:C から A:T への遷移を誘発する。誘発される点突然変異の大半は 20 種の天然アミノ酸の 1 種をコードするコドン形成する。単一点突然変異の約 12.5% しか TAG コドンを形成できないので、大きな高度突然変異誘発ゲノムライブラリーが必要である。この方法は一般に多数の新規コドンをランダムに配置した各遺伝子の突然変異コピーを少なくとも約 106 個作製する。その後、突然変異誘発前後のライブラリー個体のサブセットを配列決定することにより一般に TAG コドンが組込まれているかどうかゲノムライブラリーを確かめる。

10

20

30

40

50

【0383】

新規アミノ酸の 1 種の組込みによりどの遺伝子を改善できるかを調べるためには、場合により各種選択圧の任意のものを使用して触媒機能、タンパク質相互作用、炭素源、多重応答遺伝子及び広義代謝機能等の多様な細胞生物学をターゲットにスクリーニングを行う。例えば、トポイソメラーゼと DNA ジャイレースをターゲットとするにはキノロンに基づく選択圧を使用する。DNA 合成をターゲットとするには 5-フルオロウラシルを使用し、プロトンポンプ阻害薬をターゲットとするにはオメプラゾールを使用し、炭素利用と応答に関する各種遺伝子をターゲットとするには単一炭素源としての脂肪酸と酸性培地を使用し、チオールレドックス経路とジスルフィドを含むタンパク質をターゲットとするには還元培地を使用する。例えば Bronson, J. J. と J. F. Barrett, *Curr. Med. Chemistry*, 2001 8: p. 1775-1793; Bearden, D. T. と L. H. Danziger, *Pharmacotherapy*, 2001 21(10): p. 224S-232S; Matthews, D. A. ら, *J. Mol. Biol.* 1990. 2144(4): p. 937-948; Knox, M. R. と J. E. Harris, *Arch. Microbiol.*, 1988. 149(6): p. 557-60; McGowan, C. C., T. L. Cover, and M. J. Blaser, *Gastroenterology*, 1994. 107(5): p. 1573-8; Clark, D. P. と J. E. Cronan, *Two carbon compounds and fatty acids as carbon sources. Escheria coli and Salmonella cellular and molecular biology*, F. C. Neidhardt 編. Vol. 1. 1996, Washington D. C: ASM press; Slonczewski, J. L. と J. W. Foster, *pH-regulated genes and survival at extreme pH. Escheria coli and Salmonella cellular and molecular biology*, F. C. Neidhardt 編. Vol. 1. 1996. Wash

ington D. C. : ASM press ; 及び Ritz, D. と J. Beckwith, Annu. Rev. Microbiol. , 2001. 55 : p. 21-48 参照。

【0384】

突然変異ゲノムライブラリーをスクリーニングすると、所定選択圧下で増殖効果を付与する1組の突然変異ゲノムフラグメントが得られる。これらのフラグメントを比較し、制限マッピングとシーケンシングにより非天然アミノ酸を加えないスクリーニングと同一であるか否かを調べる。非天然アミノ酸に起因する増殖亢進を生じるフラグメントを選択し、場合によりこのフラグメントを再スクリーニングし、増殖率を各非天然アミノ酸とTAGコドンを抑圧するチロシンに比較する。例えば図21参照。選択したゲノムフラグメントのこの再スクリーニングは非天然アミノ酸が増殖効果を付与する要因であることを確証する。非天然アミノ酸をTAGコドンに挿入して選択増殖効果を示すフラグメントについては、効果を付与する遺伝子の場合により分離し、例えば消化とサブクロニングにより同定する。タンパク質を試験し、非天然アミノ酸が細胞機能をどのように亢進しているかを確認することができる。亢進したタンパク質の場合により精製し、*in vitro* 及び *in vivo* 試験の両方で天然アミノ酸と比較する。場合により標準酵素技術を使用してタンパク質安定性、動態及び他の生合成経路成分とその相互作用を検討する。

10

【実施例17】

【0385】

実施例17： 配列

20

配列番号	配列	注	tRNA又はRS
1		M. jannaschii mtRNA _{CUA} ^{Tyr}	tRNA
2		HLAD03 ; 最適化アンバーサプレッサーtRNA	tRNA
3		HL325A ; 最適化AGGAフレームシフトサプレッサーtRNA	tRNA
4		突然変異体 TyrRS (LWJ16)	RS
5		p-iPr-PheRS	RS
6		p-NH ₂ -PheRS (1)	RS
7		p-NH ₂ -PheRS (2)	RS
8		p-NH ₂ -PheRS (3a)	RS
9		p-iPr-PheRS (3b)	RS
10		O-Allyl-TyrRS (1)	RS
11		O-Allyl-TyrRS (3)	RS

配列番号	配列	注	tRNA又はRS

10

20

30

40

1 2		O-Allyl-TyrRS (4)	R S
1 3		p-Br-PheRS	R S
1 4		p-Az-PheRS (1)	R S
1 5		p-Az-PheRS (3)	R S
1 6		p-Az-PheRS (5)	R S
1 7		m-アシルフェニルアラニンをタンパク質 に取込むための突然変異体シンテターゼ (ケトン 3-4)	R S
1 8		m-アシルフェニルアラニンをタンパク質 に取込むための突然変異体シンテターゼ (ケトン 3-7)	R S
1 9		m-アシルフェニルアラニンをタンパク質 に取込むための突然変異体シンテターゼ (ケトン 4-1)	R S
2 0		m-アシルフェニルアラニンをタンパク質 に取込むための突然変異体シンテターゼ (ケトン 5-4)	R S

配列番号	配列	注	tRNA又はRS

2 1		m-アシルフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (ケトン6-8)	R S
2 2		m-メトキシフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (OMe 1-6)	R S
2 3		m-メトキシフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (OMe 1-8)	R S
2 4		m-メトキシフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (OMe 2-7)	R S
2 5		m-メトキシフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (OMe 4-1)	R S
2 6		m-メトキシフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (OMe 4-8)	R S

10

20

30

配列番号	配列	注	tRNA又はRS
2 7		p-O-アリルチロシンをタンパク質に取組むための突然変異体シンテターゼ	R S

40

		(アリル)	
28		p-ベンゾイル-L-フェニルアラニンを取込むためのアミノアシル tRNA シンテターゼクローン (p-BpaRS (H6))	RS
29		p-アジドフェニルアラニンを取込むためのアミノアシル tRNA シンテターゼクローン (p-Az-PheRS (3))	RS
30		p-アジドフェニルアラニンを取込むためのアミノアシル tRNA シンテターゼクローン (p-Az-PheRS (6))	RS
31		p-アジドフェニルアラニンを取込むためのアミノアシル tRNA シンテターゼクローン (p-Az-PheRS (20))	RS
32		p-アジドフェニルアラニンを取込むためのアミノアシル tRNA シンテターゼクローン (p-Az-PheRS (24))	RS

配列番号	配列	注	tRNA又はRS
33		Archaeoglobus fulgidus ロイシル tRNA シンテターゼ (AFLRS)	RS
34		Methanobacterium thermoautorophicum ロ	RS

		イシル tRNA シンテターゼ (MtLRS)	
3 5		突然変異体 TyrRS (LWJ16)	R S
3 6		TyrRS (SS12)	R S
3 7		p-iPr-PheRS	R S
3 8		p-NH ₂ -PheRS (1)	R S
3 9		p-NH ₂ -PheRS (2)	R S
4 0		p-NH ₂ -PheRS (3a)	R S
4 1		p-NH ₂ -PheRS (3b)	R S
4 2		0-アリル-TyrRS (1)	R S
4 3		0-アリル-TyrRS (3)	R S
4 4		0-アリル-TyrRS (3)	R S

10

20

30

配列番号	配列	注	tRNA又はRS
4 5		p-Br-PheRS	R S
4 6		p-Az-PheRS (1)	R S

40

47		p-Az-PheRS (3)	RS
48		p-Az-PheRS (5)	RS
49		m-アシルフェニルアラニンをタンパク質 に取込むための突然変異体シンテターゼ (ケトン3-4)	RS
50		m-アシルフェニルアラニンをタンパク質 に取込むための突然変異体シンテターゼ (ケトン3-7)	RS
51		m-アシルフェニルアラニンをタンパク質 に取込むための突然変異体シンテターゼ (ケトン4-1)	RS
52		m-アシルフェニルアラニンをタンパク質 に取込むための突然変異体シンテターゼ (ケトン5-4)	RS

10

20

配列番号	配列	注	tRNA又はRS
53		m-アシルフェニルアラニンをタンパク質 に取込むための突然変異体シンテターゼ (ケトン6-8)	RS
54		m-メトキシフェニルアラニンをタンパク 質に取込むための突然変異体シンテターゼ	RS

30

40

		(OMe 1-6)	
55		m-メトキシフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (OMe 1-8)	RS
56		m-メトキシフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (OMe 2-7)	RS
57		m-メトキシフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (OMe 4-1)	RS
58		m-メトキシフェニルアラニンをタンパク質に取込むための突然変異体シンテターゼ (OMe 4-8)	RS

10

20

配列番号	配列	注	tRNA又はRS
59		p-オーアシルチロシンをタンパク質に取組むための突然変異体シンテターゼ (アシル)	RS
60		p-ベンゾイル-L-フェニルアラニンを取込むためのアミノアシル tRNA シンテシンテターゼクローン (p-BpaRS (H6))	RS

30

40

6 1		p-アジドフェニルアラニンを取込むためのアミノアシル tRNA シンテターゼクロールン(p-Az-PheRS(3))	R S
6 2		p-アジドフェニルアラニンを取込むためのアミノアシル tRNA シンテターゼクロールン(p-Az-PheRS(6))	R S
6 3		p-アジドフェニルアラニンを取込むためのアミノアシル tRNA シンテターゼクロールン(p-Az-PheRS(20))	R S
6 4		p-アジドフェニルアラニンを取込むためのアミノアシル tRNA シンテターゼクロールン(p-Az-PheRS(24))	R S

10

20

配列番号	配列	注	tRNA又はRS
6 5		Archaeoglobus fulgidus ロイシル tRNA シンテターゼ(AFLRS)	R S
6 6		Methanobacterium thermoautorophicum ロイシル tRNA シンテターゼ(MtLRS)	R S
6 7		(plasc-papabc)	プラスミド
6 8		3 遺伝子(papABC)	プラスミド

30

40

【0386】

以上、分かり易くする目的で本発明を詳細に説明したが、以上の開示から当業者に自明の通り、本発明の真の範囲を逸脱することなく形態や細部に種々の変更を加えることができる。例えば、上記全技術及び装置は種々に組合せて使用することができる。本明細書に引用した全刊行物、特許、特許出願及び／又は他の文献は全目的でその開示内容全体を参考資料として組込み、個々の刊行物、特許、特許出願及び／又は他の文献を全目的で参考資料として組込むと示しているものとして扱う。

50

【図面の簡単な説明】

【0387】

【図1】 *TyrRS* の活性部位におけるアミノ酸残基の立体図である。

【図2】 各種条件下における野生型 (wt) 及び突然変異体両者の大腸菌 *DHFR* タンパク質の蓄積を示し、図2Aは精製 *DHFR* の銀染色 SDS-PAGEゲルであり、図2Bは図2Aのゲルのウェスタンブロットである。

【図3】 *DHFR* の NH₂ 末端ペプチド *MIY*MI A A L A V D R* のタンデム質量スペクトルを示す。

【図4】 各種条件下における野生型 (wt) 及び突然変異体両者のマウス *DHFR* タンパク質の蓄積を示し、図4Aは精製 *DHFR* の銀染色 SDS-PAGEゲルであり、図4Bは図4Aのゲルのウェスタンブロットである。 10

【図5】 トリプシン消化ペプチド *LLPEX*TGVLSEVQEEK* (式中、X* は L-3-(2-ナフチル)アラニンを表す) のタンデム質量スペクトルを示す。

【図6A】 増幅性蛍光レポーター系の特徴を示す。図6Aはプラスミド *pREP* である。

【図6B】 増幅性蛍光レポーター系の特徴を示す。図6Bは *pREP* (1~12) 内の *T7RNA* ポリメラーゼ遺伝子構築物の組成と蛍光増加を示す。

【図6C】 増幅性蛍光レポーター系の特徴を示す。図6Cは *pREP* (10) と *pQD* (上段) 又は *pQ* (下段) を含む細胞のサイトメトリー分析を示す。

【図6D】 増幅性蛍光レポーター系の特徴を示す。図6Dは *pREP* (10) を含み、各種大腸菌サプレッサー *tRNA* を発現する細胞の蛍光分析を示す。 20

【図7A】 *M. jannaschii TyrRS* の進化を誘導するための多目的レポータープラスミド系の成分を示す。図7Aはプラスミド *pREP/YC-JYCUA* を示す。

【図7B】 *M. jannaschii TyrRS* の進化を誘導する為の多目的レポータープラスミド系の成分を示す。図7Bは *M. jannaschii TyrRS* の進化のターゲットとして使用した非天然アミノ酸の構造を示す。

【図7C】 *M. jannaschii TyrRS* の進化を誘導するための多目的レポータープラスミド系の成分を示す。図7Cはプラスミド *pREP/YC-JYCUA* を使用してアミノアシル-*tRNA* シンテターゼを進化させるためのストラテジーを示す。

【図8A】 成功した各進化実験からの優性シンテターゼ変異体の活性を示す。図8Aは *pREP/YC-JYCUA* と指定シンテターゼ変異体を含む細胞を対応する非天然アミノ酸の存在下 (+) 又は不在下 (-) で増殖させ、長波長紫外線を照射した場合の写真である。 30

【図8B】 成功した各進化実験からの優性シンテターゼ変異体の活性を示す。図8Bは *pREP/YC-JYCUA* と指定シンテターゼ変異体を含む細胞を対応する非天然アミノ酸の存在下 (左) 又は不在下 (右) で増殖させた場合の蛍光分析を示す。

【図8C】 成功した各進化実験からの優性シンテターゼ変異体の活性を示す。図8Cは *pREP/YC-JYCUA* と指定シンテターゼ変異体を含む細胞を対応する非天然アミノ酸の存在下又は不在下で増殖させた場合の *Cm IC50* 分析を示す表である。

【図8D】 成功した各進化実験からの優性シンテターゼ変異体の活性を示す。図8Dは *pBAD/JYAMB-4TAG* と指定シンテターゼ変異体を含む細胞を対応する非天然アミノ酸の存在下 (+) 又は不在下 (-) で増殖させた場合のタンパク質発現分析を示す。 40

【図9】 ネガティブ *FACS* スクリーニング [*OAY-RS* (1, 3, 5)] 又はネガティブバルナーゼ選択 [*OAY-RS* (B)] を使用して得られた *OAS-RS* 変異体の活性比較を示す。

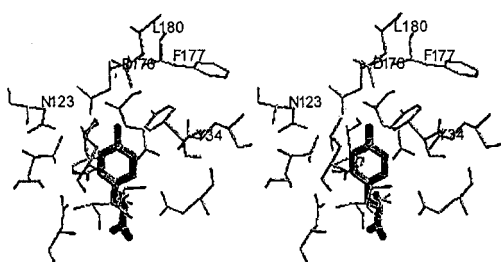
【図10】 *m-MeO-Phe* と *m-アセチル-Phe* を組込んだ *DHFR* の発現を示すウェスタンブロットのオートラジオグラフ。

【図11】 フルオレセインヒドラジドで標識したタンパク質の蛍光スペクトルを示す。

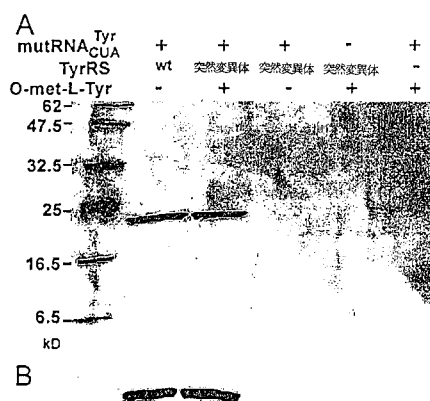
【図12】 非天然アミノ酸パラアジドフェニルアラニン及びパラベンゾイルフェニルアラニンを示す。 50

- 【図 1 3】 アリル置換フェニルアラニンの合成の為の化学スキームを示す。
- 【図 1 4】 メタ置換フェニルアラニンの合成のための化学スキームを示す。
- 【図 1 5】 p-アミノフェニルアラニンの生合成を示す。
- 【図 1 6】 各種非天然アミノ酸を示す。
- 【図 1 7】 各種非天然アミノ酸を示す。
- 【図 1 8】 各種非天然アミノ酸を示す。
- 【図 1 9】 *in vivo* 抑圧によりタンパク質に組込む付加的天然及び非天然アミノ酸を示す。
- 【図 2 0】 dopa の生産のための生合成スキームを示す。
- 【図 2 1】 21 種のアミノ酸を特異的に組込むことができる結果として細胞に得られる進 10
化的効果を測定するための方法を示す。
- 【図 2 2】 非天然アミノ酸の部位特異的組込み方法を示す。
- 【図 2 3】 各種グルタミン類似体の合成を示す。
- 【図 2 4】 γ 置換グルタミン類似体の合成を示す。
- 【図 2 5】 環状グルタミン誘導体の合成を示す。
- 【図 2 6】 各種チロシン類似体を示す。
- 【図 2 7】 チロシン類似体の生産のための合成スキームを示す。
- 【図 2 8】 グリコシル化アミノ酸を生産するための生合成スキームを示す。
- 【図 2 9】 例えば細胞取込み試験で使用する各種非天然アミノ酸を示す。

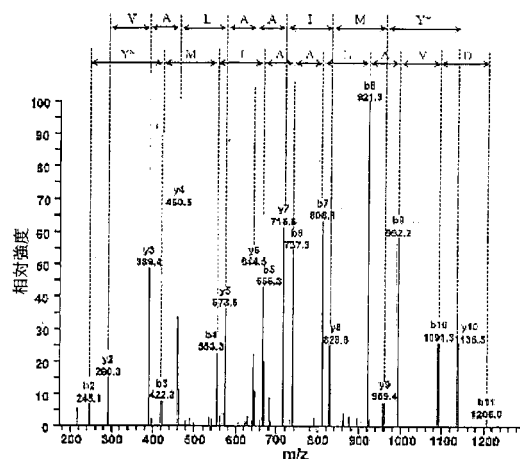
【図 1】



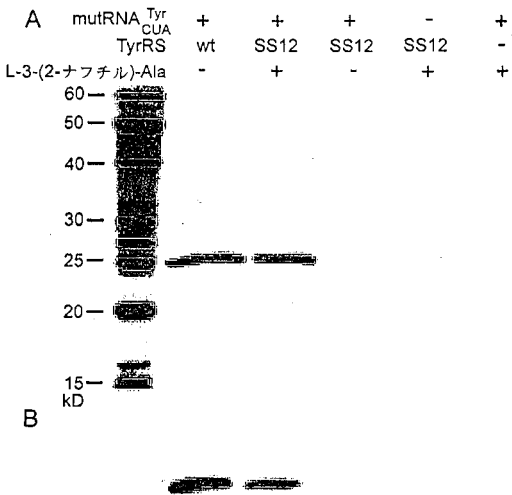
【図 2】



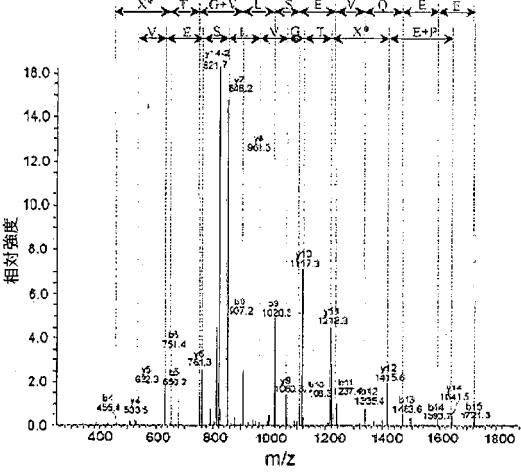
【図 3】



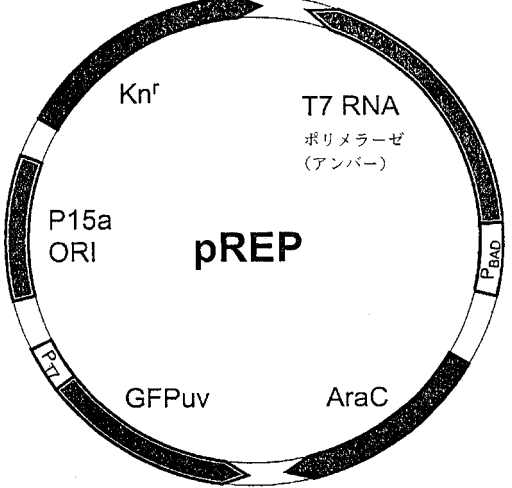
【図 4】



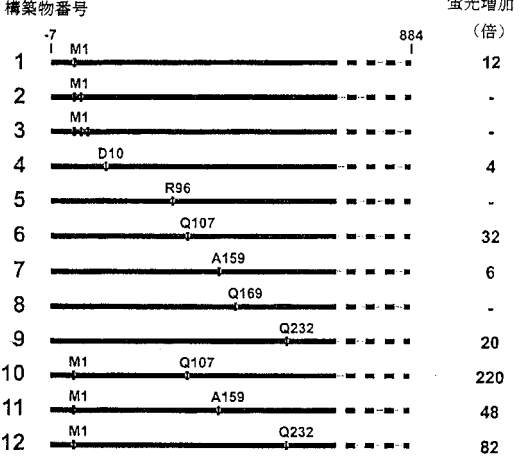
【図 5】



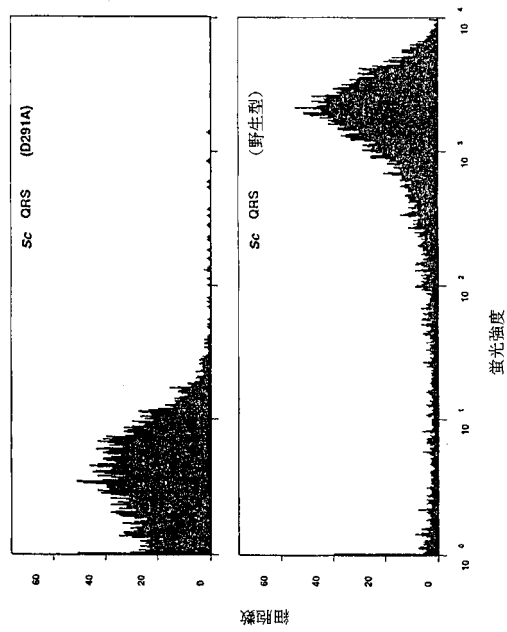
【図 6 A】



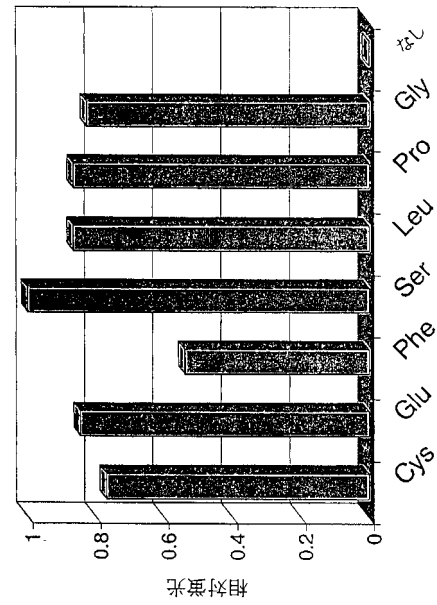
【図 6 B】



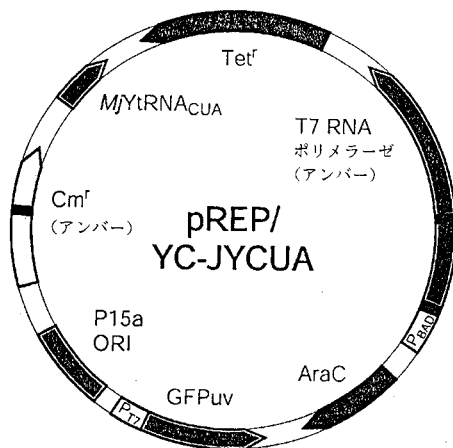
【図 6 C】



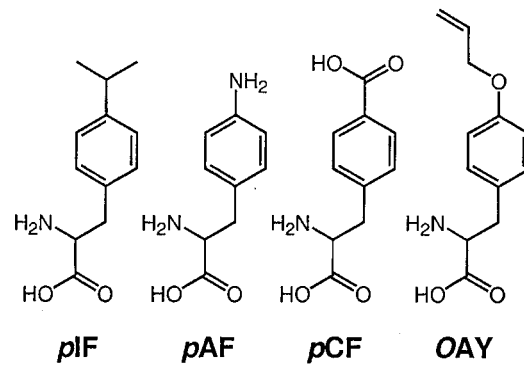
【図 6 D】



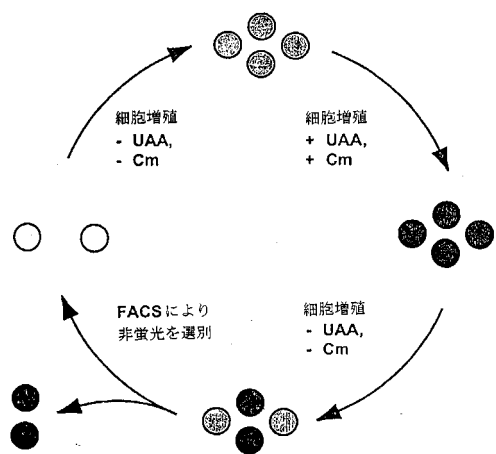
【図 7 A】



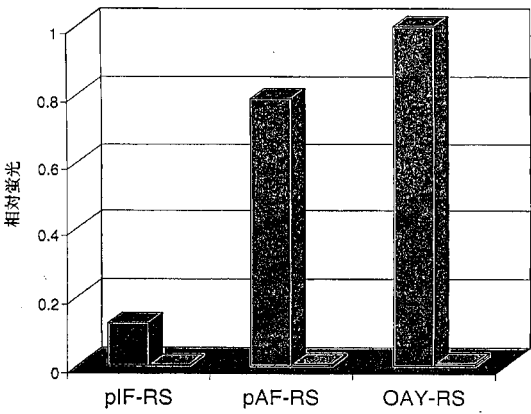
【図 7 B】



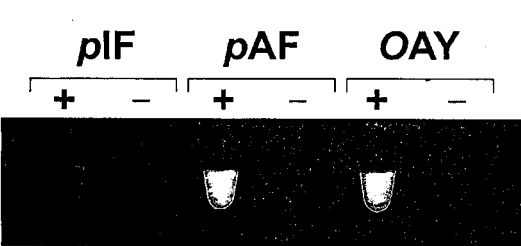
【図 7 C】



【図 8 B】



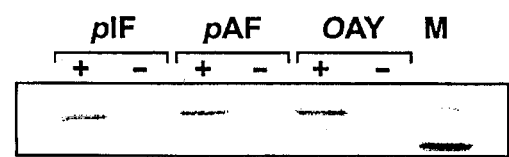
【図 8 A】



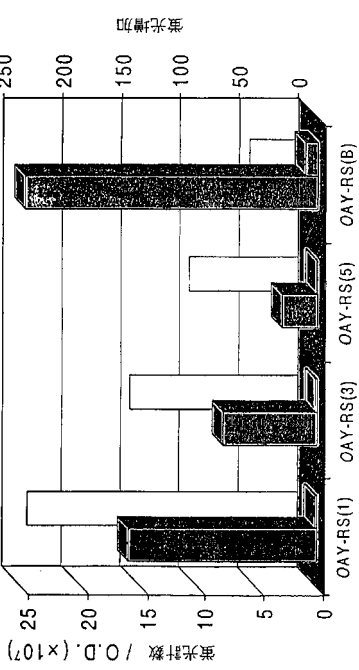
【図 8 C】

Cm IC ₅₀ (μg/mL)	pIF-RS	pAF-RS	OAY-RS(1)
- 非天然アミノ酸	< 5	< 5	< 5
+ 非天然アミノ酸	75	100	120

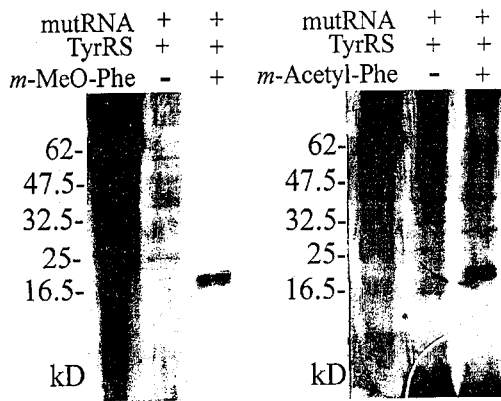
【図 8 D】



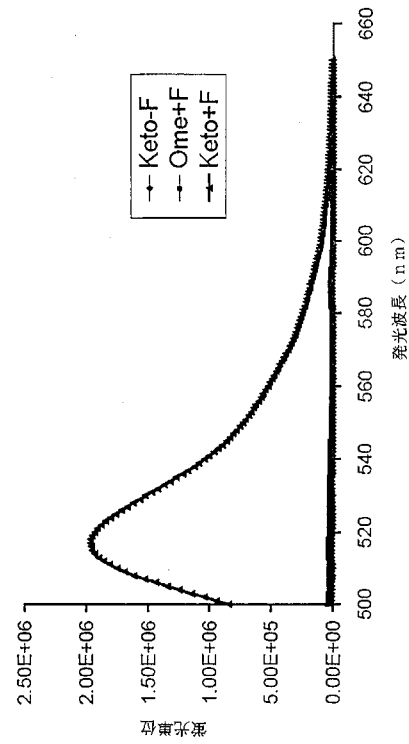
【図 9】



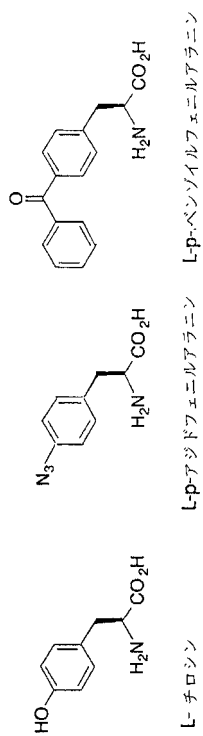
【図 10】



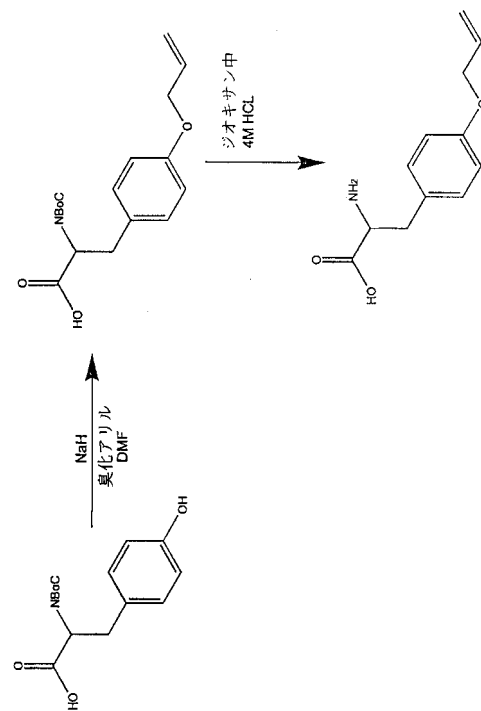
【図 11】



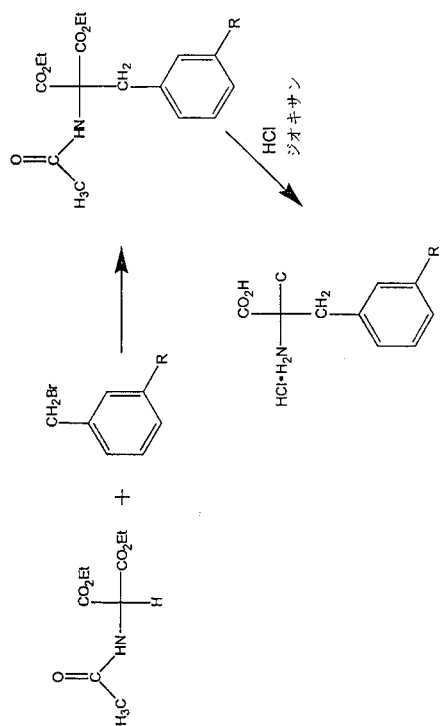
【図 12】



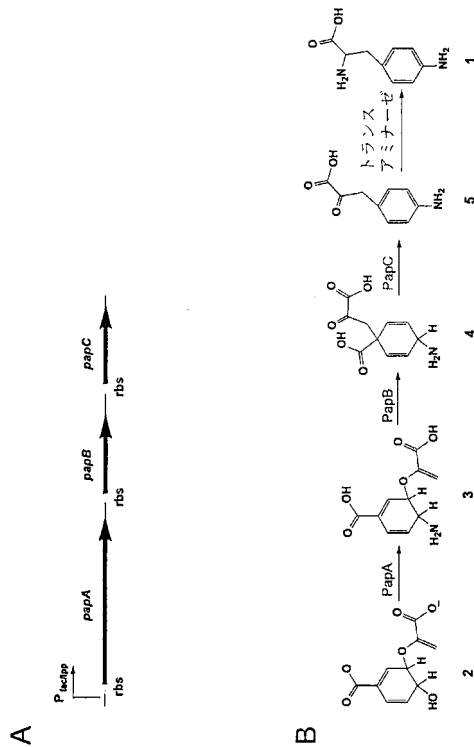
【図 13】



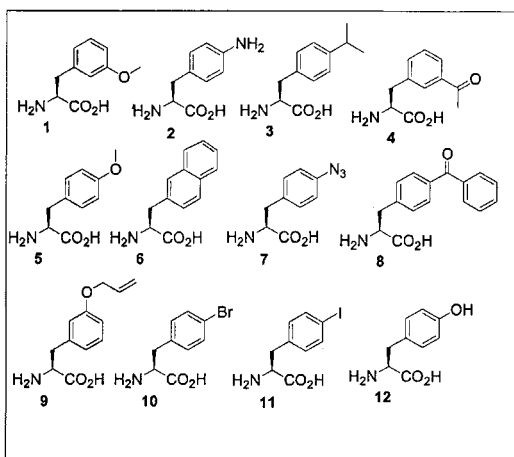
【図 14】



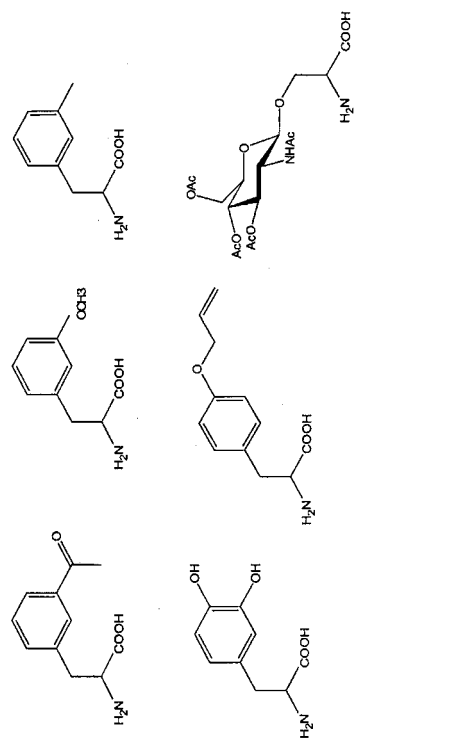
【図 15】



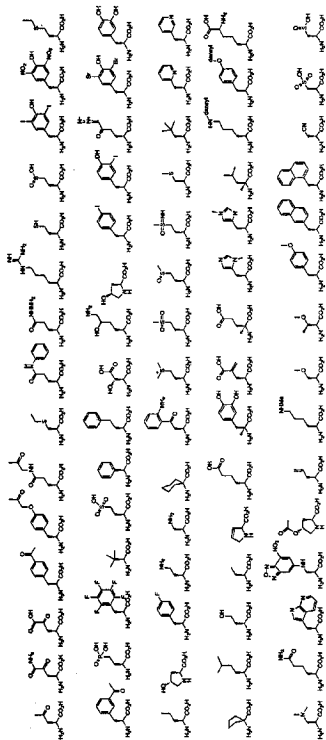
【図 16】



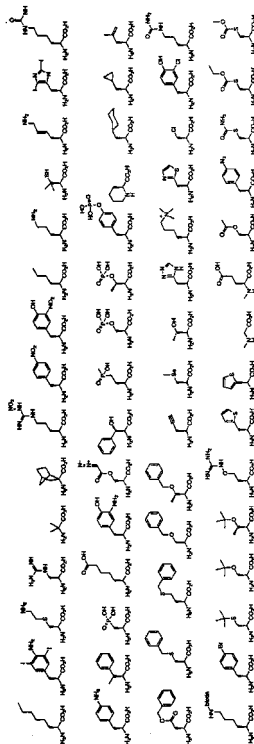
【図 17】



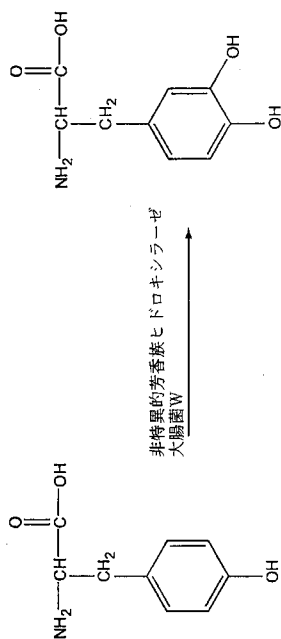
【図 18】



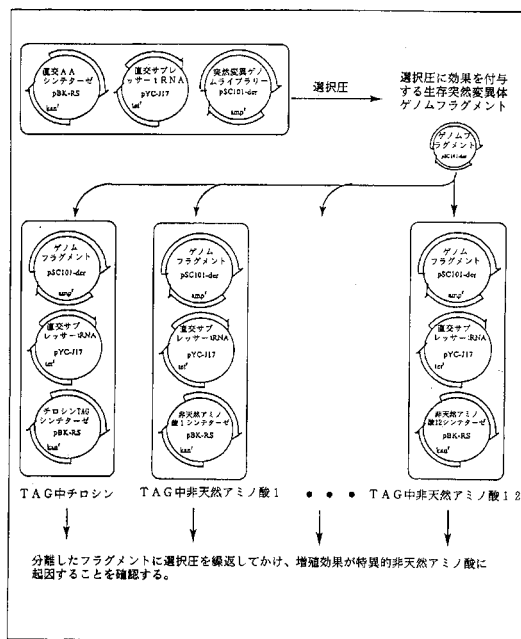
【図 19】



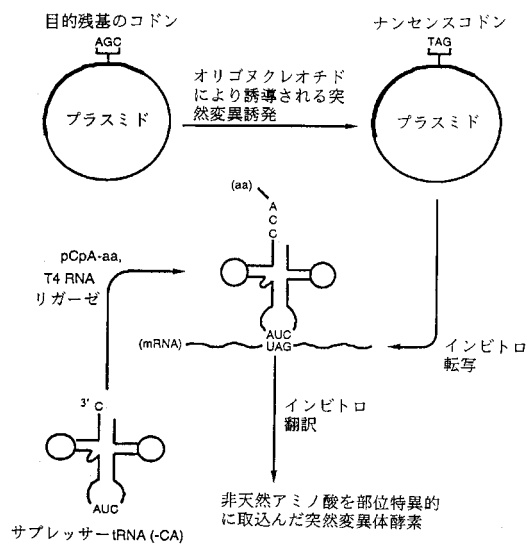
【図 20】



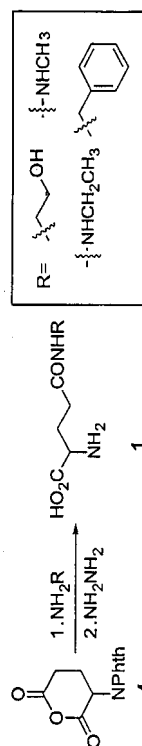
【図 21】



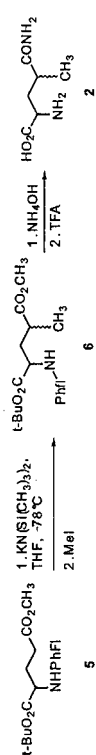
【図 2 2】



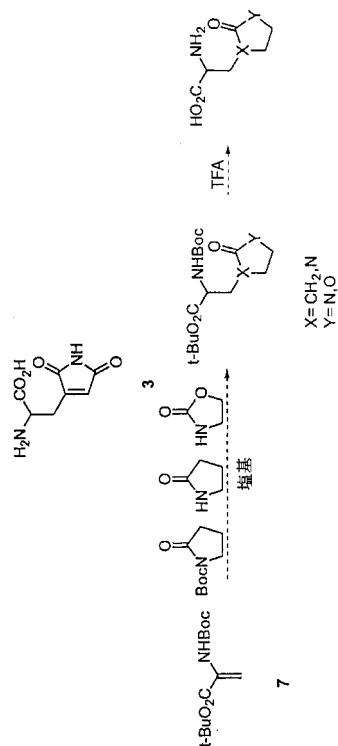
【図 2 3】



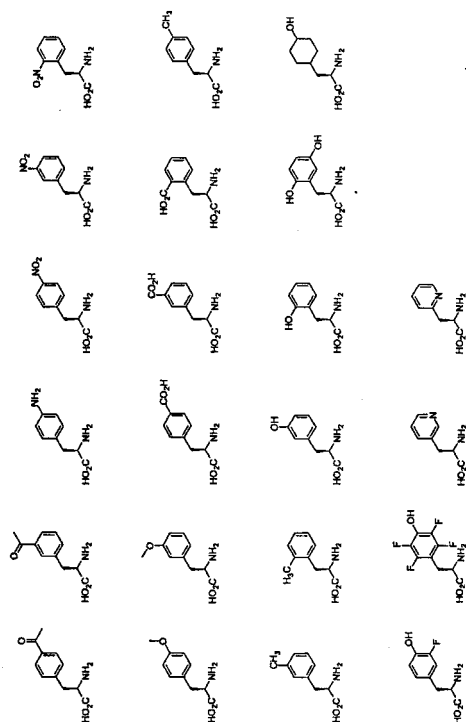
【図 2 4】



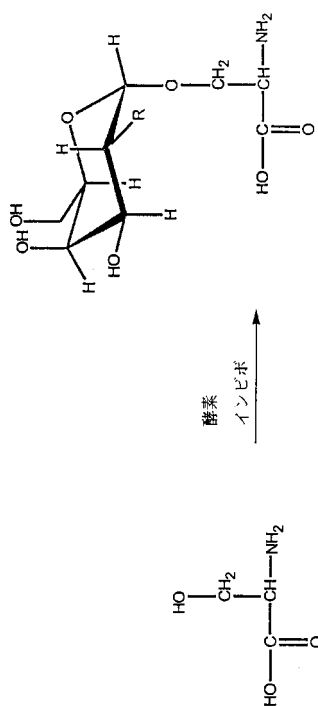
【図 2 5】



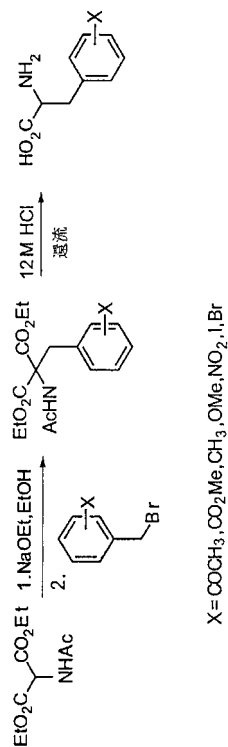
【図 26】



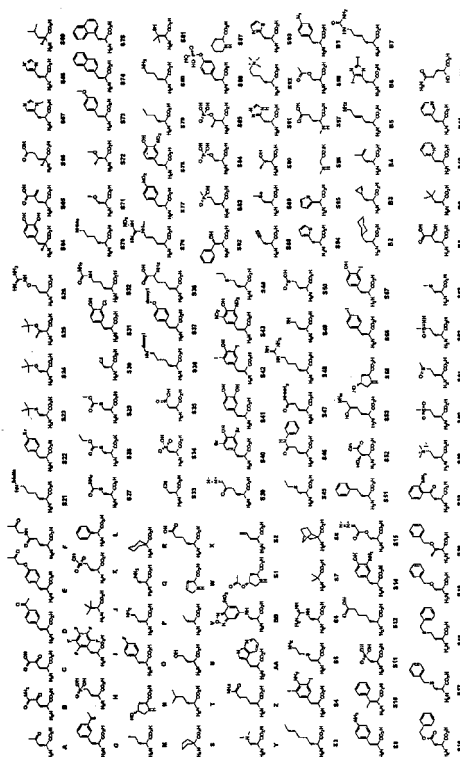
【図 28】



【図 27】



【図 29】



【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
31 October 2002 (31.10.2002)

PCT

(10) International Publication Number
WO 02/085923 A2

- (51) International Patent Classification: C07K 92130 (US). SANTORO, Steven, William [US/US]; 12668-63 Carmel Country Rd., San Diego, CA 92130 (US). ZHANG, Zhiwen [CN/US]; 7045 Charmant Drive, #33, San Diego, CA 92122 (US).
- (21) International Application Number: PCT/US02/12465
- (22) International Filing Date: 19 April 2002 (19.04.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/285,030 19 April 2001 (19.04.2001) US
60/255,514 6 February 2002 (06.02.2002) US
- (71) Applicant (for all designated States except US): THE SCRIPPS RESEARCH INSTITUTE [US/US]; 10550 North Torrey Pines Road, La Jolla, CA 92037 (US).
- (72) Inventors; and
(75) Inventors/Applicants (for US only): SCHULTZ, Peter [US/US]; 1650 La Jolla Ranch Road, La Jolla, CA 92037 (US). WANG, Lei [CN/US]; # 2320, 8730 Costa Verde Blvd., San Diego, CA 92122 (US). ANDERSON, John, Christopher [US/US]; 3922 8th Avenue, Apt. 11, San Diego, CA 92103 (US). CHIN, Jason, W., K. [GB/US]; 4249 Nobel Drive #23, San Diego, CA 92122 (US). LIU, David, R. [US/US]; 1 Fox Run Lane, Lexington, MA 02420 (US). MAGLIERY, Thomas, J. [US/US]; 333 Mill Rd., North Haven, CT 06473 (US). MEGGERS, Eric, L. [DE/US]; 8609 Via Mallorca, #13, La Jolla, CA 92037 (US). MEHL, Ryan, Aaron [US/US]; 3576 31st Street, San Diego, CA 92104 (US). PASTRNAK, Miro [SK/US]; 11686 Carmel Creek Road, Apt. 107, San Diego, CA 92130 (US).
- (81) Designated States (national): AI, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, HE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/085923 A2

(54) Title: IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS

(57) Abstract: The invention provides methods and compositions for in vivo incorporation of unnatural amino acids. Also provided are compositions including proteins with unnatural amino acids.

WO 02/085923

PCT/US02/12465

IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS

CROSS-REFERENCES TO RELATED APPLICATIONS

- 5 [01] This application claims priority to and benefit of United States provisional patent application Serial No. 60/285,030, filed April 19, 2001, and United States provisional patent application Serial No. 60/355,514, filed February 6, 2002, the specifications of which are incorporated herein in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

- 10 [02] The invention was made with United States Government support under Grant No. 6502573 from the Office of Naval Research and Grant No. GM62159 from the National Institutes of Health. The United States Government has certain rights in the invention.

FIELD OF THE INVENTION

- 15 [03] The invention relates to the field of protein biochemistry. In particular, the invention relates to the field of compositions and methods for producing proteins that include unnatural amino acids.

BACKGROUND OF THE INVENTION

- 20 [04] Proteins carry out virtually all of the complex processes of life, from photosynthesis to signal transduction and the immune response. To understand and control these intricate activities, a better understanding of the relationship between the structure and function of proteins is needed.
- [05] Unlike small organic molecule synthesis wherein almost any structural change can be made to influence functional properties of a compound, the synthesis of proteins is limited to changes encoded by the twenty natural amino acids. The genetic code of every known organism, from bacteria to human, encodes the same twenty common amino acids. These amino acids can be modified by posttranslational modification of proteins, e.g., glycosylation, phosphorylation or oxidation, or in rarer instances, by the enzymatic modification of aminoacylated suppressor tRNAs, e.g., in the case of selenocysteine.
- 30 Nonetheless, polypeptides, which are synthesized from only these 20 simple building blocks, carry out all of the complex processes of life.

WO 02/085923

PCT/US02/12465

- [06] Both site-directed and random mutagenesis, in which specific amino acids in a protein can be replaced with any of the other nineteen common amino acids, have become important tools for understanding the relationship between the structure and function of proteins. These methodologies have made possible the generation of proteins with enhanced properties, including stability, catalytic activity and binding specificity.
- 5 Nevertheless, changes in proteins are limited to the 20 common amino acids, most of which have simple functional groups. See Knowles, J.R. *Tinkering with enzymes: what are we learning?* Science, 236(4806) 1252-1258 (1987); and, Zoller, M.J., Smith, M. *Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors*, Methods Enzymol. 154:468-500 (1983). By expanding the genetic code to include additional amino acids with novel biological, chemical or physical properties, the properties of proteins, e.g., the size, acidity, nucleophilicity, hydrogen-bonding, hydrophobic properties, can be modified as compared to a protein composed of only amino acids from the 20 common amino acids, e.g., as in a naturally occurring protein.
- 10 [07] Several strategies have been employed to introduce unnatural amino acids into proteins. The first experiments involved the derivatization of amino acids with reactive side-chains such as Lys, Cys and Tyr, for example, the conversion of lysine to N²-acetyl-lysine. Chemical synthesis also provides a straightforward method to incorporate unnatural amino acids, but routine solid-phase peptide synthesis is generally limited to small peptides or proteins with less than 100 residues. With the recent development of enzymatic ligation and native chemical ligation of peptide fragments, it is possible to make larger proteins, but the method is not easily scaled. See, e.g., P. E. Dawson and S. B. H. Kent, Annu. Rev. Biochem., 69:923 (2000). A general in vitro biosynthetic method in which a suppressor tRNA chemically acylated with the desired unnatural amino acid is
- 25 added to an in vitro extract capable of supporting protein biosynthesis, has been used to site-specifically incorporate over 100 unnatural amino acids into a variety of proteins of virtually any size. See, e.g., V. W. Cornish, D. Mendel and P. G. Schultz, Angew. Chem. Int. Ed. Engl., 1995, 34:621 (1995); C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, *A general method for site-specific incorporation of unnatural amino acids into proteins*, Science 244 182-188 (1989); and, J.D. Bain, C.G. Glabe, T.A. Dix, A.R. Chamberlin, E.S. Dala, *Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide*, J. Am. Chem. Soc. 111 8013-8014 (1989). A broad range of functional groups has been introduced into proteins for studies of protein stability, protein folding, enzyme mechanism, and signal transduction. Although these studies demonstrate
- 30

WO 02/085923

PCT/US02/12465

that the protein biosynthetic machinery tolerates a wide variety of amino acid side chains, the method is technically demanding, and yields of mutant proteins are low.

- [08] Over 50 years ago, it was found that many analogs of natural amino acids inhibit the growth of bacteria. Analysis of the proteins produced in the presence of these amino acid analogs revealed that they had been substituted for their natural counterparts, to various extents. *See, e.g.,* M. H. Richmond, Bacteriol. Rev., 26:398 (1962). This occurs because the aminoacyl-tRNA synthetase, the enzyme responsible for the attachment of the correct amino acid to its cognate tRNA, cannot rigorously distinguish the analog from the corresponding natural amino acid. For instance, norleucine is charged by methionyl-tRNA synthetase, and p-fluorophenylalanine is charged by phenylalanine-tRNA synthetase. *See, D. B. Cowie, G. N. Cohen, E. T. Bolton and H. DeRoobincho-Szulmajst, Biochim. Biophys. Acta, 1959, 34:39 (1959); and, R. Munier and G. N. Cohen, Biochim. Biophys. Acta, 1959, 31:378 (1959).*
- [09] An in vivo method, termed selective pressure incorporation, was later developed to exploit the promiscuity of wild-type synthetases. *See, e.g.,* N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. M. Dong, L. Moroder and R. Huber, FASEB J., 13:41 (1999). An auxotrophic strain, in which the relevant metabolic pathway supplying the cell with a particular natural amino acid is switched off, is grown in minimal media containing limited concentrations of the natural amino acid, while transcription of the target gene is repressed. At the onset of a stationary growth phase, the natural amino acid is depleted and replaced with the unnatural amino acid analog. Induction of expression of the recombinant protein results in the accumulation of a protein containing the unnatural analog. For example, using this strategy, o, m and p-fluorophenylalanines have been incorporated into proteins, and exhibit two characteristic shoulders in the UV spectrum which can be easily identified, *see, e.g.,* C. Minks, R. Huber, L. Moroder and N. Budisa, Anal. Biochem., 284:29 (2000); trifluoromethionine has been used to replace methionine in bacteriophage T4 lysozyme to study its interaction with chitooligosaccharide ligands by ¹⁹F NMR, *see, e.g.,* H. Duwel, E. Daub, V. Robinson and J. F. Honek, Biochemistry, 36:3404 (1997); and trifluoroleucine has been inserted in place of leucine, resulting in increased thermal and chemical stability of a leucine-zipper protein. *See, e.g.,* Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado and D. A. Tirrell, Angew. Chem. Int. Ed. Engl., 40:1494 (2001). Moreover, selenomethionine and telluromethionine are incorporated into various recombinant proteins to facilitate the solution of phases in X-ray crystallography. *See, e.g.,* W. A. Hendrickson, J. R. Horton and D. M. Lemaster, EMBO

WO 02/085923

PCT/US02/12465

- J., 9:1665 (1990); J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odom, B. Dunlap, L. Lebioda and M. Hatada, Nat. Struct. Biol., 1:283 (1994); N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann and R. Huber, Eur. J. Biochem., 230:788 (1995); and, N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neuefeind, L.
- 5 Moroder and R. Huber, J. Mol. Biol., 270:616 (1997). Methionine analogs with alkene or alkyne functionalities have also been inserted efficiently, allowing for additional modification of proteins by chemical means. See, e.g., J. C. M. van Hest and D. A. Tirrell, FEBS Lett., 428:68 (1998); J. C. M. van Hest, K. L. Kiick and D. A. Tirrell, J. Am. Chem. Soc., 122:1282 (2000); and, K. L. Kiick and D. A. Tirrell, Tetrahedron, 56:9487 (2000).
- 10 [10] The success of this method depends on the recognition of the unnatural amino acid analogs by aminoacyl-tRNA synthetases, which, in general, require high selectivity to insure the fidelity of protein translation. Therefore, the range of chemical functionality accessible via this route is limited. For instance, although thiaproline can be incorporated quantitatively into proteins, oxaproline and selenoproline cannot. See, N. Budisa, C.
- 15 Minks, F. J. Medrano, J. Lutz, R. Huber and L. Moroder, Proc. Natl. Acad. Sci. U S A, 95:455 (1998). One way to expand the scope of this method is to relax the substrate specificity of aminoacyl-tRNA synthetases, which has been achieved in a limited number of cases. For example, it was found that replacement of Ala²⁹⁴ by Gly in *Escherichia coli* phenylalanyl-tRNA synthetase (PheRS) increases the size of substrate binding pocket, and
- 20 results in the acylation of tRNA^{Phe} by p-Cl-phenylalanine (p-Cl-Phe). See, M. Ibba, P. Kast and H. Hennecke, Biochemistry, 33:7107 (1994). An *Escherichia coli* strain harboring this mutant PheRS allows the incorporation of p-Cl-phenylalanine or p-Br-phenylalanine in place of phenylalanine. See, e.g., M. Ibba and H. Hennecke, FEBS Lett., 364:272 (1995); and, N. Sharma, R. Furter, P. Kast and D. A. Tirrell, FEBS Lett., 467:37
- 25 (2000). Similarly, a point mutation Phe130Ser near the amino acid binding site of *Escherichia coli* tyrosyl-tRNA synthetase was shown to allow azatyrosine to be incorporated more efficiently than tyrosine. See, F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll and S. Nishimura, J. Biol. Chem., 275:40324 (2000).
- 30 [11] The fidelity of aminoacylation is maintained both at the level of substrate discrimination and proofreading of non-cognate intermediates and products. Therefore, an alternative strategy to incorporate unnatural amino acids into proteins in vivo is to modify synthetases that have proofreading mechanisms. These synthetases cannot discriminate and therefore activate amino acids that are structurally similar to the cognate natural amino

WO 02/085923

PCT/US02/12465

acids. This error is corrected at a separate site, which deacylates the mischarged amino acid from the tRNA to maintain the fidelity of protein translation. If the proofreading activity of the synthetase is disabled, structural analogs that are misactivated may escape the editing function and be incorporated. This approach has been demonstrated recently with the valyl-tRNA synthetase (ValRS). See, V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel and P. Marliere, *Science*, 292:501 (2001). ValRS can misaminoacylate tRNA^{Val} with Cys, Thr, or aminobutyrate (Abu); these noncognate amino acids are subsequently hydrolyzed by the editing domain. After random mutagenesis of the *Escherichia coli* chromosome, a mutant *Escherichia coli* strain was selected that has a mutation in the editing site of ValRS. This edit-defective ValRS incorrectly charges tRNA^{Val} with Cys. Because Abu sterically resembles Cys (–SH group of Cys is replaced with –CH₃ in Abu), the mutant ValRS also incorporates Abu into proteins when this mutant *Escherichia coli* strain is grown in the presence of Abu. Mass spectrometric analysis shows that about 24% of valines are replaced by Abu at each valine position in the native protein.

[12] At least one major limitation of the methods described above is that all sites corresponding to a particular natural amino acid throughout the protein are replaced. The extent of incorporation of the natural and unnatural amino acid may also vary – only in rare cases can quantitative substitution be achieved since it is difficult to completely deplete the cognate natural amino acid inside the cell. Another limitation is that these strategies make it difficult to study the mutant protein in living cells, because the multisite incorporation of analogs often results in toxicity. Finally, this method is applicable in general only to close structural analogs of the common amino acids, again because substitutions must be tolerated at all sites in the genome.

[13] Solid-phase synthesis and semisynthetic methods have also allowed for the synthesis of a number of small proteins containing novel amino acids. For example, see the following publications and references cited within, which are as follows: Crick, F.J.C., Barrett, L. Brenner, S. Watts-Tobin, R. *General nature of the genetic code for proteins. Nature*, 1227-1232 (1961); Hofmann, K., Bohn, H. *Studies on polypeptides. XXXVI. The effect of pyrazole-imidazole replacements on the S-protein activating potency of an S-peptide fragment, J. Am Chem*, 5914-5919 (1966); Kaiser, E.T. *Synthetic approaches to biologically active peptides and proteins including enzymes, Acc Chem Res*, 47-54 (1989); Nakatsuka, T., Sasaki, T., Kaiser, E.T. *Peptide segment coupling catalyzed by the semisynthetic enzyme thiosubtilisin, J Am Chem Soc*, 3808-3810 (1987); Schnolzer, M.,

WO 02/085923

PCT/US02/12465

Kent, S B H. *Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease*, Science, 221-225 (1992); Chaiken, I.M. *Semisynthetic peptides and proteins*, CRC Crit Rev Biochem, 255-301 (1981); Offord, R.E. *Protein engineering by chemical means?* Protein Eng., 151-157 (1987); and, Jackson, D.Y.,

- 5 Burnier, J., Quan, C., Stanley, M., Tom, J., Wells, J.A. *A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues*, Science, 243 (1994).
- [14] Chemical modification has been used to introduce a variety of unnatural side chains, including cofactors, spin labels and oligonucleotides into proteins in vitro. *See, e.g.,* Corey, D.R., Schultz, P.G. *Generation of a hybrid sequence-specific single-stranded deoxyribonuclease*, Science, 1401-1403 (1987); Kaiser, E.T., Lawrence D.S., Rokita, S.E. *The chemical modification of enzymatic specificity*, Rev Biochem, 565-595 (1985); Kaiser, E.T., Lawrence, D.S. *Chemical mutation of enzyme active sites*, Science, 505-511 (1984); Neet, K.E., Nanci A, Koshland, D.E. *Properties of thiol-subtilisin*, J Biol. Chem, 15 6392-6401 (1968); Polgar, L.B., M.L. *A new enzyme containing a synthetically formed active site. Thiol-subtilisin*, J. Am Chem Soc, 3153-3154 (1966); and, Pollack, S.J., Nakayama, G. Schultz, P.G. *Introduction of nucleophiles and spectroscopic probes into antibody combining sites*, Science, 1038-1040 (1988).
- [15] Alternatively, biosynthetic methods that employ chemically modified aminoacyl-
20 tRNAs have been used to incorporate several biophysical probes into proteins synthesized in vitro. See the following publications and references cited within: Brunner, J. *New Photolabeling and crosslinking methods*, Annu. Rev Biochem, 483-514 (1993); and, Krieg, U.C., Walter, P., Hohnson, A.E. *Photocrosslinking of the signal sequence of nascent preprolactin of the 54-kilodalton polypeptide of the signal recognition particle*,
25 Proc. Natl. Acad. Sci, 8604-8608 (1986).
- [16] Previously, it has been shown that unnatural amino acids can be site-specifically incorporated into proteins in vitro by the addition of chemically aminoacylated suppressor tRNAs to protein synthesis reactions programmed with a gene containing a desired amber nonsense mutation. Using these approaches, one can substitute a number of the common
30 twenty amino acids with close structural homologues, e.g., fluorophenylalanine for phenylalanine, using strains auxotrophic for a particular amino acid. *See, e.g.,* Noren, C.J., Anthony-Cahill, Griffith, M.C., Schultz, P.G. *A general method for site-specific incorporation of unnatural amino acids into proteins*, Science, 244: 182-188 (1989); M.W. Nowak, et al., Science 268:439-42 (1995); Bain, J.D., Glabe, C.G., Dix, T.A.,

WO 02/085923

PCT/US02/12465

- Chamberlin, A.R., Dimala, E.S. *Biosynthetic site-specific Incorporation of a non-natural amino acid into a polypeptide*, J. Am Chem Soc., 111:8013-8014 (1989); N. Budisa et al., FASEB J. 13:41-51 (1999); Ellman, J.A., Mendel, D., Anthony-Cahill, S., Noren, C.J., Schultz, P.G. *Biosynthetic method for introducing unnatural amino acids site-specifically into proteins*, Methods in Enz., 301-336 (1992); and, Mendel, D., Cornish, V.W. & Schultz, P.G. *Site-Directed Mutagenesis with an Expanded Genetic Code*, Annu Rev Biophys. Biomol Struct. 24, 435-62 (1995).
- [17] For example, a suppressor tRNA was prepared that recognized the stop codon UAG and was chemically aminoacylated with an unnatural amino acid. Conventional site-directed mutagenesis was used to introduce the stop codon TAG, at the site of interest in the protein gene. See, e.g., Sayers, J.R., Schmidt, W. Eckstein, F. 5', 3' *Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucleic Acids Res. 791-802 (1988). When the acylated suppressor tRNA and the mutant gene were combined in an in vitro transcription/translation system, the unnatural amino acid was incorporated in response to the UAG codon which gave a protein containing that amino acid at the specified position. Experiments using [³H]-Phe and experiments with α -hydroxy acids demonstrated that only the desired amino acid is incorporated at the position specified by the UAG codon and that this amino acid is not incorporated at any other site in the protein. See, e.g., Noren, et al, *supra*; and, Ellman, J.A., Mendel, D., Schultz, P.G. *Site-specific incorporation of novel backbone structures into proteins*, Science, 197-200 (1992).
- [18] In general, these in vitro approaches are limited by difficulties in achieving site-specific incorporation of the amino acids, by the requirement that the amino acids be simple derivatives of the common twenty amino acids or problems inherent in the synthesis of large proteins or peptide fragments.
- [19] Microinjection techniques have also been use incorporate unnatural amino acids into proteins. See, e.g., M. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. G. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty and H. A. Lester, Science, 268:439 (1995); and, D. A. Dougherty, Curr. Opin. Chem. Biol., 4:645 (2000). A *Xenopus* oocyte was coinjected with two RNA species made in vitro: an mRNA encoding the target protein with a UAG stop codon at the amino acid position of interest and an amber suppressor tRNA aminoacylated with the desired unnatural amino acid. The translational machinery of the oocyte then inserts the unnatural amino acid at the position specified by UAG. This

WO 02/085923

PCT/US02/12465

method has allowed in vivo structure-function studies of integral membrane proteins, which are generally not amenable to in vitro expression systems. Examples include the incorporation of a fluorescent amino acid into tachykinin neurokinin-2 receptor to measure distances by fluorescence resonance energy transfer, see, e.g., G. Turcatti, K. Nemeth, M.

- 5 D. Edgerton, U. Meseth, F. Talabot, M. Peitsch, J. Knowles, H. Vogel and A. Chollet, J. Biol. Chem., 271:19991 (1996); the incorporation of biotinylated amino acids to identify surface-exposed residues in ion channels, see, e.g., J. P. Gallivan, H. A. Lester and D. A. Dougherty, Chem. Biol., 4:739 (1997); the use of caged tyrosine analogs to monitor conformational changes in an ion channel in real time, see, e.g., J. C. Miller, S. K.
- 10 Silverman, P. M. England, D. A. Dougherty and H. A. Lester, Neuron, 20:619 (1998); and, the use of alpha hydroxy amino acids to change ion channel backbones for probing their gating mechanisms. See, e.g., P. M. England, Y. Zhang, D. A. Dougherty and H. A. Lester, Cell, 96:89 (1999); and, T. Lu, A. Y. Ting, J. Mainland, L. Y. Jan, P. G. Schultz and J. Yang, Nat. Neurosci., 4:239 (2001).

- 15 [20] However, there are limitations microinjection method, e.g., the suppressor tRNA has to be chemically aminoacylated with the unnatural amino acid in vitro, and the acylated tRNA is consumed as a stoichiometric reagent during translation and cannot be regenerated. This limitation results in poor suppression efficiency and low protein yields, necessitating highly sensitive techniques to assay the mutant protein such as
- 20 electrophysiological measurements. Moreover, this method is only applicable to cells that can be microinjected.

- [21] The ability to incorporate unnatural amino acids directly into proteins in vivo offers the advantages of high yields of mutant proteins, technical ease, the potential to study the mutant proteins in cells or possibly in living organisms and the use of these
- 25 mutant proteins in therapeutic treatments. The ability to include unnatural amino acids with various sizes, acidities, nucleophilicities, hydrophobicities, and other properties into proteins can greatly expand our ability to rationally and systematically manipulate the structures of proteins, both to probe protein function and create new proteins or organisms with novel properties. However, the process is difficult, because the complex nature of
- 30 tRNA-synthetase interactions that are required to achieve a high degree of fidelity in protein translation.

- [22] In one attempt to site-specifically incorporate para-F-Phe, a yeast amber suppressor tRNA^{Phe}CUA /phenylalanyl-tRNA synthetase pair was used in a p-F-Phe resistant, Phe auxotrophic *Escherichia coli* strain. See, e.g., R. Furter, Protein Sci., 7:419 (1998).

WO 02/085923

PCT/US02/12465

Because yeast PheRS does not have high substrate specificity for p-F-Phe, the mutagenesis site was translated with only 64-75% p-F-Phe and the remainder as Phe and Lys even in the excess of p-F-Phe added to the growth media. In addition, at the Phe codon positions, 7% p-F-Phe was found, indicating that the endogenous *Escherichia coli* PheRS

5 incorporates p-F-Phe in addition to Phe. Besides of its translational infidelity, e.g., the suppressor tRNA and PheRS are not truly orthogonal, this approach is not generally applicable to other unnatural amino acids.

[23] Therefore, improvements to the process are needed to provide more efficient and effective methods to alter the biosynthetic machinery of the cell. The present invention
10 addresses these and other needs, as will be apparent upon review of the following disclosure.

SUMMARY OF THE INVENTION

[24] The present invention provides a variety of methods for making and using translation systems that can incorporate unnatural amino acids into proteins, as well as
15 related compositions. Proteins comprising unnatural amino acids made by the translation system are also a feature of the invention. Both known and new unnatural amino acids can be incorporated into proteins using the translation system of the invention. The invention further provides novel unnatural amino acids; various compositions including the unnatural amino acids, e.g., proteins and cells including unnatural amino acids; chemical
20 and biosynthetic methods for producing unnatural amino acids; and methods for producing and compositions comprising an autonomous twenty-one amino acid cell.

[25] Thus, in one aspect, the present invention provides compositions comprising a translation system. The translation system comprises an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl tRNA synthetase (O-RS). Typically, the O-RS preferentially
25 aminoacylates the O-tRNA with at least one unnatural amino acid in the translation system and the O-tRNA recognizes at least one selector codon. The translation system thus inserts the unnatural amino acid into a protein produced in the system, in response to an encoded selector codon.

[26] Typical translation systems include cells, such as bacterial cells (e.g., *Escherichia coli*), archaeobacterial cells, eukaryotic cells (e.g., yeast cells, mammalian cells, plant
30 cells, insect cells), or the like. Alternatively, the translation system comprises an in vitro translation system, e.g., a translation extract including a cellular extract.

WO 02/085923

PCT/US02/12465

[27] Example O-tRNAs comprise a nucleic acid comprising a polynucleotide sequence selected from the group consisting of: SEQ ID NO:1-3 and/or a complementary polynucleotide sequence thereof. Similarly, example O-RS include polypeptides selected from the group consisting of: a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 35-66 and a polypeptide encoded by a nucleic acid comprising a polynucleotide sequence selected from the group consisting of: SEQ ID NO: 4-34 and a complementary polynucleotide sequence thereof.

[28] Examples of unnatural amino acids that can be used by the translation system include: an unnatural analogue of a tyrosine amino acid; an unnatural analogue of a glutamine amino acid; an unnatural analogue of a phenylalanine amino acid; an unnatural analogue of a serine amino acid; an unnatural analogue of a threonine amino acid; an alkyl, aryl, acyl, azido, cyano, halo, hydrazine, hydrazide, hydroxyl, alkenyl, alkynyl, ether, thiol, sulfonyl, seleno, ester, thioacid, borate, boronate, phospho, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, hydroxylamine, keto, or amino substituted amino acid, or any combination thereof; an amino acid with a photoactivatable cross-linker; a spin-labeled amino acid; a fluorescent amino acid; an amino acid with a novel functional group; an amino acid that covalently or noncovalently interacts with another molecule; a metal binding amino acid; a metal-containing amino acid; a radioactive amino acid; a photocaged and/or photoisomerizable amino acid; a biotin or biotin-analogue containing amino acid; a glycosylated or carbohydrate modified amino acid; a keto containing amino acid; amino acids comprising polyethylene glycol or polyether; a heavy atom substituted amino acid; a chemically cleavable or photocleavable amino acid; an amino acid with an elongated side chain; an amino acid containing a toxic group; a sugar substituted amino acid, e.g., a sugar substituted serine or the like; a carbon-linked sugar-containing amino acid; a redox-active amino acid; an α -hydroxy containing acid; an amino thio acid containing amino acid; an α,α disubstituted amino acid; a β -amino acid; and a cyclic amino acid other than proline.

[29] For example, the unnatural amino acid can be an O-methyl-L-tyrosine, an L-3-(2-naphthyl)alanine, a 3-methyl-phenylalanine, an O-4-allyl-L-tyrosine, a 4-propyl-L-tyrosine, a tri-O-acetyl-GlcNAc β -serine, an L-Dopa, a fluorinated phenylalanine, an isopropyl-L-phenylalanine, a *p*-azido-L-phenylalanine, a *p*-acyl-L-phenylalanine, a *p*-benzoyl-L-phenylalanine, an L-phosphoserine, a phosphoserine, a phosphotyrosine, a *p*-iodo-phenylalanine, a *p*-bromophenylalanine, a *p*-amino-L-phenylalanine, and an isopropyl-L-

WO 02/085923

PCT/US02/12465

phenylalanine. In one embodiment, the at least one unnatural amino acid is an O-methyl-L-tyrosine. In one specific example embodiment, the at least one unnatural amino acid is an L-3-(2-naphthyl)alanine. In another set of specific examples, the at least one unnatural amino acid is an amino-, isopropyl-, or O-allyl-containing phenylalanine analogue.

5 [30] Any of a variety of selector codons can be used in the present invention, including nonsense codons, rare codons, four (or more) base codons, or the like. For example, in one embodiment, the at least one selector codon is an amber codon.

[31] A variety of exemplar translation systems are provided herein, including e.g., an Escherichia coli cell comprising a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ and a mutant TyrRS (LWJ16), where the mutant TyrRS (LWJ16) preferentially aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with O-methyl-L-tyrosine in the cell and the cell uses the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ to recognize an amber codon. In another example, an Escherichia coli cell comprising a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ and an SS12-TyrRS is provided, where the SS12-TyrRS preferentially aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with L-3-(2-naphthyl)alanine in the cell and the cell uses the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ to recognize an amber

15 codon.

[32] The translation system herein provides the ability to synthesize proteins that comprise unnatural amino acids in usefully large quantities. For example, proteins comprising at least one unnatural amino acid can be produced at a concentration of at least about 10, 50, 100 or more micrograms per liter, e.g., in a composition comprising a cell

20 extract, a buffer, a pharmaceutically acceptable excipient, and/or the like.

[33] Another aspect of the present invention provides for the production of proteins that are homologous to any available protein, but comprising one or more unnatural amino acid homologue. For example, therapeutic proteins can be made that comprise one or more unnatural amino acid and are homologous to one or more therapeutic protein. For

25 example, in one aspect, the protein is homologous to a therapeutic or other protein such as: a cytokine, a growth factor, a growth factor receptor, an interferon, an interleukin, an inflammatory molecule, an oncogene product, a peptide hormone, a signal transduction molecule, a steroid hormone receptor, a transcriptional activator, a transcriptional suppressor, erythropoietin (EPO), insulin, human growth hormone, epithelial Neutrophil

30 Activating Peptide-78, GRO α /MGSA, GRO β , GRO γ , MIP-1 α , MIP-1 β , MCP-1, hepatocyte growth factor, insulin-like growth factor, leukemia inhibitory factor, oncostatin M, PD-ECSF, PDGF, pleiotropin, SCF, c-kit ligand, VEGF, G-CSF, IL-1, IL-2, IL-8,

WO 02/085923

PCT/US02/12465

IGF-I, IGF-II, FGF (fibroblast growth factor), PDGF, TNF, TGF- α , TGF- β , EGF (epidermal growth factor), KGF (keratinocyte growth factor), SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, hyalurin/CD44, Mos, Ras, Raf, Met; p53, Tat, Fos, Myc, Jun, Myb, Rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and/or corticosterone. In another set of embodiments, the protein is homologous to a therapeutic or other protein such as: an Alpha-1 antitrypsin, an Angiostatin, an Antihemolytic factor, an antibody, an Apolipoprotein, an Apoprotein, an Atrial natriuretic factor, an Atrial natriuretic polypeptide, an Atrial peptide, a C-X-C chemokine, T39765, NAP-2, ENA-78, a Gro-a, a Gro-b, a Gro-c, an IP-10, a GCP-2, an NAP-4, an SDF-1, a PF4, a MIG, a Calcitonin, a c-kit ligand, a cytokine, a CC chemokine, a Monocyte chemoattractant protein-1, a Monocyte chemoattractant protein-2, a Monocyte chemoattractant protein-3, a Monocyte inflammatory protein-1 alpha, a Monocyte inflammatory protein-1 beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262, a CD40, a CD40 ligand, a C-kit Ligand, a Collagen, a Colony stimulating factor (CSF), a Complement factor 5a, a Complement inhibitor, a Complement receptor 1, a cytokine, an epithelial Neutrophil Activating Peptide-78, a GRO α /MGSA, a GRO β , a GRO γ , a MIP-1 α , a MIP-16, a MCP-1, an Epidermal Growth Factor (EGF), an epithelial Neutrophil Activating Peptide, an Erythropoietin (EPO), an Exfoliating toxin, a Factor IX, a Factor VII, a Factor VIII, a Factor X, a Fibroblast Growth Factor (FGF), a Fibrinogen, a Fibronectin, a G-CSF, a GM-CSF, a Glucocerebrosidase, a Gonadotropin, a growth factor, a growth factor receptor, a Hedgehog protein, a Hemoglobin, a Hepatocyte Growth Factor (HGF), a Hirudin, a Human serum albumin, an ICAM-1, an ICAM-1 receptor, an LFA-1, an LFA-1 receptor, an Insulin, an Insulin-like Growth Factor (IGF), an IGF-I, an IGF-II, an interferon, an IFN- α , an IFN- β , an IFN- γ , an interleukin, an IL-1, an IL-2, an IL-3, an IL-4, an IL-5, an IL-6, an IL-7, an IL-8, an IL-9, an IL-10, an IL-11, an IL-12, a Keratinocyte Growth Factor (KGF), a Lactoferrin, a leukemia inhibitory factor, a Luciferase, a Neurturin, a Neutrophil inhibitory factor (NIF), an oncostatin M, an Osteogenic protein, an oncogene product, a Parathyroid hormone, a PD-ECSF, a PDGF, a peptide hormone, a Human Growth Hormone, a Pleiotropin, a Protein A, a Protein G, a Pyrogenic exotoxins A, B, or C, a Relaxin, a Renin, an SCF, a Soluble complement receptor I, a Soluble I-CAM 1, a Soluble interleukin receptors, a Soluble TNF receptor, a Somatomedin, a Somatostatin, a Somatotropin, a Streptokinase, a Superantigens, a Staphylococcal enterotoxins, an SEA, an SEB, an SEC1, an SEC2, an SEC3, an SED, an

WO 02/085923

PCT/US02/12465

- SEE, a steroid hormone receptor, a Superoxide dismutase, a Toxic shock syndrome toxin, a Thymosin alpha 1, a Tissue plasminogen activator, a tumor growth factor (TGF), a TGF- α , a TGF- β , a Tumor Necrosis Factor, a Tumor Necrosis Factor alpha, a Tumor necrosis factor beta, a Tumor necrosis factor receptor (TNFR), a VLA-4 protein, a VCAM-1
- 5 protein, a Vascular Endothelial Growth Factor (VEGF), a Urokinase, a Mos, a Ras, a Raf, a Met, a p53, a Tat, a Fos, a Myc, a Jun, a Myb, a Rel, an estrogen receptor, a progesterone receptor, a testosterone receptor, an aldosterone receptor, an LDL receptor, and/or a corticosterone. In one aspect, the compositions herein comprise a protein comprising an unnatural amino acid and a pharmaceutically acceptable excipient, including,
- 10 e.g., any of the proteins noted above and a pharmaceutically acceptable excipient.
- [34] Homology to the polypeptide can be inferred by performing a sequence alignment, e.g., using BLASTN or BLASTP, e.g., set to default parameters. For example, in one embodiment, the protein is at least about 50%, at least about 75%, at least about 80%, at least about 90% or at least about 95% identical to a known therapeutic protein (e.g., a
- 15 protein present in Genbank or other available databases). For example, in one preferred embodiment, the therapeutic protein is erythropoietin (EPO).
- [35] The protein of interest can contain 1, 2, 3, 4, 5, 6, 7, 6, 9, 10, 11, 12, 13, 14, 15 or more unnatural amino acids. The unnatural amino acids can be the same or different, e.g., there can be 1, 2, 3, 4, 5, 6, 7, 6, 9, 10, 11, 12, 13, 14, 15 or more different sites in the
- 20 protein that comprise 1, 2, 3, 4, 5, 6, 7, 6, 9, 10, 11, 12, 13, 14, 15 or more different unnatural amino acids. For example, in one embodiment, the protein is DHFR, and the at least one unnatural amino acid is selected from the group consisting of O-methyl-L-tyrosine and L-3-(2-naphthyl)alanine.
- [36] The present invention also provides methods for producing at least one protein in
- 25 a translation system such that the at least one protein comprises at least one unnatural amino acid. In the methods, the translation system is provided with at least one nucleic acid comprising at least one selector codon, wherein the nucleic acid encodes the at least one protein. The translation system is also provided with an orthogonal tRNA (O-tRNA), that functions in the translation system and recognizes the at least one selector codon and
- 30 an orthogonal aminoacyl tRNA synthetase (O-RS), that preferentially aminoacylates the O-tRNA with the at least one unnatural amino acid in the translation system. The translation system is also provided with the at least one unnatural amino, thereby

WO 02/085923

PCT/US02/12465

producing, in the translation system, the at least one protein comprising the at least one unnatural amino acid.

[37] All of the above structural features of the compositions can be embodied in the methods, e.g., types of translation systems (e.g., cells, cell extracts, etc.), types of proteins produced in the translation systems (e.g., EPO homologues and the other proteins noted
5 herein) specific mutant proteins, specific unnatural amino acids, and the like.

[38] In one aspect, the protein(s) comprising unnatural amino acids that are produced are processed and modified in a cell-dependent manner. This provides for the production of proteins that are stably folded, glycosylated, or otherwise modified by the cell.

10 [39] The unnatural amino acid is optionally provided exogenously to the translation system. Alternately, e.g., where the translation system is a cell, the unnatural amino acid can be biosynthesized by the translation system.

[40] In one specific example embodiment, the invention provides methods for producing in an *Escherichia coli* cell at least one protein comprising at least one O-methyl-
15 L-tyrosine. The method includes providing the translation system with at least one nucleic acid comprising an amber codon, wherein the nucleic acid encodes the at least one protein; providing the translation system with a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$, wherein the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ functions in the cell and wherein the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ recognizes the amber codon; providing the translation system with a mutant TyrRS (LWJ16), wherein the mutant TyrRS (LWJ16) aminoacylates
20 the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with the O-methyl-L-tyrosine in the cell; and, providing the cell with the O-methyl-L-tyrosine, thereby producing in the cell at least one protein comprising the O-methyl-L-tyrosine.

[41] In another example embodiment, the invention provides a method for producing in an *Escherichia coli* cell at least one protein comprising at least one L-3-(2-naphthyl)alanine. In this example embodiment, the method includes: providing the
25 translation system with at least one nucleic acid comprising an amber codon, wherein the nucleic acid encodes the at least one protein; providing the cell with a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$, wherein the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ functions in the cell and wherein the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ recognizes the amber codon; providing the cell with an SS12-TyrRS, wherein the SS12-TyrRS
30 aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with the L-3-(2-naphthyl)alanine in the cell; and, providing the cell with the L-3-(2-naphthyl)alanine, thereby producing in the cell at least one protein comprising the L-3-(2-naphthyl)alanine.

WO 02/085923

PCT/US02/12465

[42] In another aspect, the present invention provides unnatural amino acids, e.g., meta substituted phenylalanine analogues, such as 3-acetyl-phenylalanine and 3-methoxy phenylalanine; tyrosine analogues, such as 4-allyl tyrosine; glycosylated amino acids, and the like.

- 5 [43] Various compositions comprising unnatural amino acids, e.g., proteins and cells comprising the unnatural amino acids of the invention, are also provided. For example, compositions comprising an unnatural amino acid and an orthogonal tRNA, e.g., covalently bonded, are provided. Compositions comprising unnatural amino acids and an orthogonal aminoacyl tRNA synthetase, e.g., hydrogen bonded, are also provided.

- 10 [44] In another aspect, the present invention provides methods of synthesizing amino acids. For example, 4-allyl-L-tyrosine, is typically synthesized by reacting a protected tyrosine with allyl bromide, e.g., in the presence of sodium hydride and DMF, and deprotecting to yield 4-allyl-L-tyrosine. Typically an NBoc or Fmoc protected tyrosine is used, e.g., with an acidic deprotection, e.g., in the presence of hydrochloric acid and dioxane. The final product is optionally extracted, e.g., with ethanol or dichloromethane.

- 15 [45] Meta-substituted phenylalanine analogues are typically synthesized by condensing diethylacetamidomalonate and a meta-substituted benzyl bromide. The product of the condensation is then typically hydrolyzed to yield the meta-substituted phenylalanine analogue, e.g., a keto, acetyl, or methoxy substituted phenylalanine such as 3-methoxy-phenylalanine or 3-acetyl-phenylalanine. The desired meta substituted benzyl bromide is optionally synthesized by reacting N-bromosuccinimide (NBS) with 3-methylacetophenone to produce a brominated product, and crystallizing the brominated product in a hexane solution. The crystallization yields a monobromide product as opposed to a mixture of a monobromide and a dibromide.

- 25 [46] In another aspect, the present invention provides biosynthetic methods for producing unnatural amino acids. For example, glycosylated amino acids are optionally synthesized in vivo, e.g., by transforming a cell with a plasmid comprising a gene for an N-acetyl-galactosaminidase, a transglycosylase, or a serine-glycosylhydrolase. The cell then produces the desired glycosylated amino acid, e.g. from cellular resources. In another example, *p*-aminophenylalanine is synthesized, e.g., in vivo, by enzymatically converting chorismate to 4-amino-4-deoxychorismic acid; which is enzymatically converted to 4-amino-4-deoxyphenic acid; and enzymatically converting the 4-amino-4-deoxyphenic acid to *p*-aminophenyl-pyruvic acid, which is enzymatically converted to *p*-aminophenylalanine. The enzymatic conversions are typically performed using a 4-

WO 02/085923

PCT/US02/12465

amino-4-deoxychorismate synthase, e.g., PapA, a chorismate mutase, e.g., PapB, and a prephenate dehydrogenase, e.g., PapC, respectively. The final step is typically performed by contacting the p-aminophenyl-pyruvic acid with an aminotransferase, e.g., a non-specific tyrosine aminotransferase, e.g., derived from *E. coli*. Aminotransferases of use in the present invention include, but are not limited to, tyrB, aspS, or ilvE. Typically the above steps are performed *in vivo*, e.g., by transforming a cell with a plasmid comprising the genes which encode the enzymes used for the synthesis.

[47] In another aspect, the present invention provides a method of producing p-aminophenylalanine in an *Escherichia coli* cell. The method typically comprises transforming the cell with a plasmid comprising papA, papB, and papC, wherein the cell comprises chorismate and an aminotransferase. Expression of papA, papB, and papC results in a synthase, a mutase, and a dehydrogenase, wherein these enzymes together with the aminotransferase produce p-phenylalanine from chorismate.

[48] In another aspect, the present invention provides an autonomous twenty-one (or more) amino acid cell. The cell, e.g., a bacterial cell, typically comprises a biosynthetic pathway system for producing an unnatural amino acid, e.g., p-aminophenylalanine, from one or more carbon sources within the cell, e.g., chorismate, and a translation system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl tRNA synthetase (O-RS). The O-RS preferentially aminoacylates the O-tRNA with the unnatural amino acid and the O-tRNA incorporates the unnatural amino acid into a protein in response to a selector codon, e.g., a nonsense codon such as TAG, a four base codon, or an amber codon. The cell can comprise more than one unnatural amino acid, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more unnatural amino acids, optionally with more than one orthogonal tRNA (e.g., one per unnatural amino acid to provide for site-specific incorporation of each unnatural amino acid in a protein, or more, or less, to tune the specificity of unnatural amino acid incorporation) and/or more than one orthogonal aminoacyl tRNA synthetase (O-RS) (e.g., one per orthogonal tRNA, or more or less to tune the specificity of unnatural amino acid incorporation).

[49] In some embodiments, the biosynthetic pathway systems produce a natural cellular amount of the unnatural amino acid, e.g., the cell produces the unnatural amino acid in an amount sufficient for protein biosynthesis, which amount does not substantially alter the concentration of natural amino acids or substantially exhaust cellular resources in the production of the unnatural amino acids.

WO 02/085923

PCT/US02/12465

- [50] In one example class of embodiments, the autonomous cell is engineered to produce *p*-aminophenylalanine from chorismate as described above. In this embodiment, the cell is engineered to produce the desired enzymes as described above, e.g., a synthase, a dehydrogenase, and a mutase derived from *Streptomyces Venezuelae* or *Streptomyces pristinaespiralis* and an aminotransferase derived from *E.coli*. For example, the cells of the invention are optionally transformed with a plasmid, e.g., low copy pSC101 derived plasmid, comprising papA, papB, and papC, wherein the plasmid further comprises an lpp promoter and a lac promoter. In some embodiments, the plasmid further comprises one or more ribosome binding sites.
- 10 [51] Other unnatural amino acids that are optionally produced by the cells of the invention include, but are not limited to, dopa, O-methyl-L-tyrosine, glycosylated amino acids, pegylated amino acids, other unnatural amino acids noted herein, and the like.
- [52] In another related aspect, the present invention provides a cell comprising one or more systems for producing at least twenty one amino acids and specifically incorporating one or more of the amino acids into one or more proteins within the cell, wherein at least
- 15 one of the incorporated amino acids comprises an unnatural amino acid.
- [53] In another aspect, the present invention provides a method of identifying an advantage provided by an unnatural amino acid which has been incorporated into one or more proteins of a cell. The method typically comprises providing a library of cells, each of which cells comprises a randomized plasmid, e.g., derived from an *E. coli* genome.
- 20 One or more of the randomized plasmids typically confers on the cells an ability to incorporate an unnatural amino acid into a protein. The library of cells is then screened to identify cells with enhanced growth, e.g., as compared to a native *E. coli* cell, thereby identifying an advantage provided by the unnatural amino acid. In some embodiments, a second screen is used to further verify that any advantage identified is due to the unnatural amino acid.
- 25 [54] Kits are an additional feature of the invention. For example, the kits can include one or more translation system as noted above (e.g., a cell, a 21 or more amino acid cell, etc.), one or more unnatural amino acid, e.g., with appropriate packaging material, containers for holding the components of the kit, instructional materials for practicing the methods herein and/or the like. Similarly, products of the translation systems (e.g., proteins such as EPO analogues comprising unnatural amino acids) can be provided in kit form, e.g., with containers for holding the components of the kit, instructional materials for practicing the methods herein and/or the like.
- 30

WO 02/085923

PCT/US02/12465

BRIEF DESCRIPTION OF THE FIGURES

- [55] **Figure 1** is a stereoview of the amino acid residues in the active site of TyrRS (modified from P. Brick, T. N. Bhat, D. M. Blow, *J. Mol. Biol.* 208, 83-98 (1988)). Residues from *B. stearothermophilus* TyrRS are shown in the figure.
- 5 [56] **Figure 2** illustrates accumulation of *E. coli* DHFR protein, both wild-type (wt) and mutant under different conditions. Expression conditions are notated at the top of each lane. The left lane is molecular weight marker. **Figure 2A** is a silver-stained SDS-PAGE gel of purified DHFR. **Figure 2B** is a Western blot of the gel in **Figure 2A**.
- [57] **Figure 3** is a tandem mass spectrum of an NH₂ terminal peptide of DHFR, 10 MIY*MIAALAVDR. The partial sequence Y*MIAALAVDR of the peptide containing the O-methyl-L-tyrosine residue (Y*) can be read from the annotated b or y ion series.
- [58] **Figure 4** illustrates accumulation of mouse DHFR protein, both wild-type (wt) and mutant, under different conditions. Expression conditions are notated at the top of each lane. The left lane is molecular weight marker. **Figure 4A** is a silver-stained SDS-PAGE 15 gel of purified DHFR. **Figure 4B** is a Western blot of the gel in **Figure 4A**.
- [59] **Figure 5** is a tandem mass spectrum of the tryptic peptide LLPEX*TGVLSVQEEK (X* represents L-3-(2-naphthyl)-alanine). The sequence can be read from the annotated b or y ion series; even so, b7 and y13 are not observed. The base peak 821.7 (100%) assigned to the doubly charged y14 ion is truncated for clarity.
- 20 [60] **Figure 6, Panels A-D**, illustrate features of the amplifiable fluorescence reporter system. **Figure 6A** is plasmid pREP. T7 RNA polymerase transcription is controlled by the *ara* promoter; protein expression depends on suppression of amber codons at varying locations within the gene. GFPuv expression is controlled by T7 RNA polymerase. Plasmid pREP is compatible for use with a ColE1 plasmid expressing an orthogonal 25 synthetase/tRNA pair. **Figure 6B** illustrates composition and fluorescence enhancement of T7 RNA polymerase gene constructs within pREP(1-12). The construct number is indicated to the left of each. Fluorescence enhancements, indicated to the right of each construct, are calculated as the cell concentration-corrected ratio of fluorescence, as measured fluorimetrically, of cells containing pREP(1-12) and pQ or pQD. The positions 30 of amber mutations within the gene are indicated above each construct. **Figure 6C** illustrates cytometric analysis of cells containing pREP(10) and either pQD (*top*) or pQ (*bottom*). **Figure 6D** illustrates fluorimetric analyses of cells containing pREP(10) and expressing various *E. coli* suppressor tRNAs. 'None' indicates that the cells contain no suppressor tRNA.

WO 02/085923

PCT/US02/12465

- [61] **Figure 7, Panels A-C**, illustrates components of a multipurpose reporter plasmid system for directing the evolution of *M. jannaschii* TyrRS. **Figure 7A** illustrates plasmid pREP/YC-JYCUA. Plasmid pREP/YC-JYCUA is compatible for use with plasmid pBK and variants. **Figure 7B** illustrates structures of unnatural amino acids used as targets for the evolution of *M. jannaschii* TyrRS. **Figure 7C** illustrates a strategy for a evolution of an aminoacyl-tRNA synthetase using plasmid pREP/YC-JYCUA. Fluorescent and non-fluorescent cells are shown in black and grey, respectively.
- [62] **Figure 8, Panels A-D**, illustrates the activity of the dominant synthetase variant from each successful evolution experiment. **Figure 8A** is a photograph illustrating long-wavelength ultraviolet illumination of cells containing pREP/YC-JYCUA and the indicated synthetase variant, grown in either the presence (+) or absence (-) of the corresponding unnatural amino acid. **Figure 8B** illustrates a fluorimetric analysis of cells containing pREP/YC-JYCUA and the indicated synthetase variant, grown in either the presence (left) or absence (right) of the corresponding unnatural amino acid. **Figure 8C** is a table that illustrates a Cm IC₅₀ analysis of cells containing pREP/YC-JYCUA and the indicated synthetase variant, grown in either the presence or absence of the corresponding unnatural amino acid. **Figure 8D** illustrates a protein expression analysis from cells containing pBAD/JYAMB-4TAG and the indicated synthetase variant, grown in either the presence (+) or absence (-) of the corresponding unnatural amino acid.
- [63] **Figure 9** illustrates activity comparisons of OAY-RS variants derived using a negative FACS-based screen [OAY-RS(1,3,5)] or negative barnase-based selection [OAY-RS(B)]. Cells containing pREP/YC-JYCUA and the indicated synthetase variant were grown in either the presence (solid block, left) or absence (solid block, right) of the corresponding unnatural amino acid and analyzed fluorimetrically. Fluorescence enhancement (bar, back) is calculated as the cell concentration-corrected ratio of fluorescence of cells grown in the presence versus the absence of unnatural amino acid.
- [64] **Figure 10** is an autoradiograph of a western blot demonstrating expression of *m*-MeO-Phe- and *m*-Acetyl-Phe- incorporated DHFR.
- [65] **Figure 11** illustrates the fluorescence emission spectra of fluorescein hydrazide labelled protein.
- [66] **Figure 12** illustrates the unnatural amino acids para-azido-phenylalanine and para-benzoyl-phenylalanine.
- [67] **Figure 13** illustrates a chemical scheme for the synthesis of an allyl-substituted phenylalanine.

WO 02/085923

PCT/US02/12465

- [68] Figure 14 illustrates a chemical scheme for the synthesis of meta-substituted phenylalanines.
- [69] Figure 15 illustrates the biosynthesis of *p*-aminophenylalanine. Panel A illustrates a plasmid used for the biosynthesis of *p*-aminophenylalanine and Panel B illustrates a biosynthetic scheme for the production of *p*-aminophenylalanine from chorismate, e.g., using the plasmid of Panel A.
- [70] Figure 16 illustrates a variety of unnatural amino acids.
- [71] Figure 17 illustrates a variety of unnatural amino acids.
- [72] Figure 18 illustrates a variety of unnatural amino acids.
- [73] Figure 19 illustrates additional amino acids, natural and unnatural for incorporation into proteins via *in vivo* suppression.
- [74] Figure 20 provides a biosynthetic scheme for production of dopa.
- [75] Figure 21 illustrates a method for determining evolutionary advantages in a cell due to the ability to specifically incorporate twenty-one amino acids.
- [76] Figure 22 illustrates a method for site-specific incorporation of unnatural amino acids.
- [77] Figure 23 illustrates the synthesis of various glutamine analogs.
- [78] Figure 24 illustrates the synthesis of a gamma substituted glutamine analog.
- [79] Figure 25 illustrates the synthesis of a cyclic glutamine derivative.
- [80] Figure 26 illustrates a variety of tyrosine analogs.
- [81] Figure 27 illustrates a synthetic scheme for the production of tyrosine analogs.
- [82] Figure 28 illustrates a biosynthetic scheme for producing glycosylated amino acids.
- [83] Figure 29 illustrates a variety of unnatural amino acids, e.g., as used in a cellular uptake study. Any or all of the above figures are schematic in nature.

DETAILED DESCRIPTION

IN GENERAL

- [84] The present invention provides compositions and methods for augmenting the protein biosynthetic machinery of a cell to accommodate additional genetically encoded amino acids using orthogonal tRNA/aminoacyl tRNA synthetase (O-tRNA/O-RS) pairs. The compositions and methods described here can be used with unnatural amino acids, e.g., providing novel spectroscopic, chemical or structural properties to proteins using any

WO 02/085923

PCT/US02/12465

of a wide array of side chains. The invention is applicable to both prokaryotic (e.g., Eubacteria, Archaeobacteria) and eukaryotic (e.g., yeast, mammalian, plant, or insect) cells. These compositions and methods are useful for the site specific incorporation of unnatural amino acids via selector codons, e.g., stop codons, four base codons, and the like. The invention also provides proteins, including unnatural amino acids, produced using the compositions or made by the methods of the invention. The ability to introduce unnatural amino acids into proteins directly in living cells provides new tools for studies of protein and cellular function and can lead to the generation of proteins with enhanced properties useful for, e.g., therapeutics.

10 **DEFINITIONS**

[85] **Homologous:** Proteins and/or protein sequences are "homologous" when they are derived, naturally or artificially, from a common ancestral protein or protein sequence. Similarly, nucleic acids and/or nucleic acid sequences are homologous when they are derived, naturally or artificially, from a common ancestral nucleic acid or nucleic acid sequence. For example, any naturally occurring nucleic acid can be modified by any available mutagenesis method to include one or more selector codon. When expressed, this mutagenized nucleic acid encodes a polypeptide comprising one or more unnatural amino acid. The mutation process can, of course, additionally alter one or more standard codon, thereby changing one or more standard amino acid in the resulting mutant protein as well. Homology is generally inferred from sequence similarity between two or more nucleic acids or proteins (or sequences thereof). The precise percentage of similarity between sequences that is useful in establishing homology varies with the nucleic acid and protein at issue, but as little as 25% sequence similarity is routinely used to establish homology. Higher levels of sequence similarity, e.g., 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or 99% or more can also be used to establish homology. Methods for determining sequence similarity percentages (e.g., BLASTP and BLASTN using default parameters) are described herein and are generally available.

[86] **Orthogonal:** As used herein, the term "orthogonal" refers to a molecule (e.g., an orthogonal tRNA (O-tRNA) and/or an orthogonal aminoacyl tRNA synthetase (O-RS)) that is used with reduced efficiency by a system of interest (e.g., a translational system, e.g., a cell). Orthogonal refers to the inability or reduced efficiency, e.g., less than 20 % efficient, less than 10 % efficient, less than 5 % efficient, or e.g., less than 1% efficient, of an orthogonal tRNA and/or orthogonal RS to function in the translation system of interest.

WO 02/085923

PCT/US02/12465

For example, an orthogonal tRNA in a translation system of interest aminoacylates any endogenous RS of a translation system of interest with reduced or even zero efficiency, when compared to aminoacylation of an endogenous tRNA by the endogenous RS. In another example, an orthogonal RS aminoacylates any endogenous tRNA in the translation system of interest with reduced or even zero efficiency, as compared to aminoacylation of the endogenous tRNA by an endogenous RS.

[87] Preferentially aminoacylates: The term "preferentially aminoacylates" refers to an efficiency of, e.g., about 70 % efficient, about 75 % efficient, about 85% efficient, about 90% efficient, about 95 % efficient, or about 99% or more efficient, at which an O-RS aminoacylates an O-tRNA with an unnatural amino acid compared to a naturally occurring tRNA or starting material used to generate the O-tRNA. The unnatural amino acid is then incorporated into a growing polypeptide chain with high fidelity, e.g., at greater than about 75% efficiency for a given selector codon, at greater than about 80% efficiency for a given selector codon, at greater than about 90% efficiency for a given selector codon, at greater than about 95% efficiency for a given selector codon, or at greater than about 99% or more efficiency for a given selector codon.

[88] Selector codon: The term "selector codon" refers to codons recognized by the O-tRNA in the translation process and not recognized by an endogenous tRNA. The O-tRNA anticodon loop recognizes the selector codon on the mRNA and incorporates its amino acid, e.g., an unnatural amino acid, at this site in the polypeptide. Selector codons can include, e.g., nonsense codons, such as, stop codons, e.g., amber, ochre, and opal codons; four or more base codons; codons derived from natural or unnatural base pairs and the like. For a given system, a selector codon can also include one of the natural three base codons, wherein the endogenous system does not use said natural three base codon, e.g., a system that is lacking a tRNA that recognizes the natural three base codon or a system wherein the natural three base codon is a rare codon.

[89] Suppressor tRNA: A suppressor tRNA is a tRNA that alters the reading of a messenger RNA (mRNA) in a given translation system. A suppressor tRNA can read through, e.g., a stop codon, a four base codon, or a rare codon.

[90] Translation system: The term "translation system" refers to the components necessary to incorporate a naturally occurring amino acid into a growing polypeptide chain (protein). Components of a translation system can include, e.g., ribosomes, tRNAs, synthetases, mRNA and the like. The components of the present invention can be added to a translation system, in vivo or in vitro. A translation system can be a cell, either

WO 02/085923

PCT/US02/12465

prokaryotic, e.g., an *E. coli* cell, or eukaryotic, e.g., a yeast, mammalian, plant, or insect cell.

[91] Unnatural amino acid: As used herein, the term "unnatural amino acid" refers to any amino acid, modified amino acid, and/or amino acid analogue that is not one of the 20 naturally occurring amino acids or seleno cysteine.

[92] Unless otherwise defined herein or below in the remainder of the specification, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs.

DISCUSSION

- 10 [93] Proteins are at the crossroads of virtually every biological process, from photosynthesis and vision to signal transduction and the immune response. These complex functions result from a polyamide based polymer consisting of twenty relatively simple building blocks arranged in a defined primary sequence.
- 15 [94] The present invention includes methods and composition for use in the site-specific incorporation of unnatural amino acids directly into proteins in vivo. Importantly, the unnatural amino acid is added to the genetic repertoire, rather than substituting for one of the common 20 amino acids. The present invention, e.g., (i) allows the site-selective or random insertion of one or more unnatural amino acids at any desired position of any protein, (ii) is applicable to both prokaryotic and eukaryotic cells, (iii) enables in vivo
- 20 studies of mutant proteins in addition to the generation of large quantities of purified mutant proteins, and (iv) is adaptable to incorporate any of a large variety of non-natural amino acids into proteins in vivo. The invention provides compositions and methods useful for in vivo site specific incorporation of unnatural amino acids. Specifically, the invention provides translation systems, e.g., cells, that include an orthogonal tRNA (O-
- 25 tRNA), an orthogonal aminoacyl tRNA synthetase (O-RS), and an unnatural amino acid, where the O-RS aminoacylates the O-tRNA with the unnatural amino acid, and the cell uses the components to incorporate the unnatural amino acid into a growing polypeptide chain.
- [95] The invention further provides methods for in vivo site-specific incorporation of
- 30 unnatural amino acids using the translation systems of the invention. The invention also provides proteins produced by the methods of the invention. The claimed proteins include unnatural amino acids.

WO 02/085923

PCT/US02/12465

- [96] The compositions and methods of the invention utilize an orthogonal tRNA (O-tRNA) aminoacyl tRNA synthetase (O-RS) pair. A wide range of pairs can be used with the following properties: the O-tRNA is preferentially aminoacylated with an unnatural amino acid by the O-RS. In addition, the orthogonal pair functions in the translation system of interest, e.g., the translation system uses the unnatural amino acid-aminoacylated O-tRNA to incorporate the unnatural amino acid into a polypeptide chain. Incorporation occurs in a site specific manner, e.g., the O-tRNA recognizes a selector codon, e.g., a stop codon, in the mRNA coding for the protein.
- [97] In one embodiment, the O-tRNA is derived from a Tyr-tRNA from a *Methanococcus jannaschii* cell. In a preferred embodiment, the O-tRNA is that referred to herein as $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$. In another embodiment, the O-tRNA includes a nucleic acid polynucleotide sequence selected from the group that includes SEQ ID NO: 1-3 or a complementary polynucleotide sequence thereof.
- [98] In some embodiments of the invention, the O-RS is derived from TyrRS from a *Methanococcus jannaschii* cell. In a preferred embodiment, the O-RS is referred to herein as mutant TyrRS (LWJ16) or SS12-TyrRS. In a further embodiment, the O-RS includes a polypeptide selected from the group consisting of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 35-66 and a polypeptide encoded by a nucleic acid comprising a polynucleotide sequence selected from the group consisting of: SEQ ID NO: 4-34 or a complementary polynucleotide sequence thereof.
- [99] In a preferred embodiment, the invention includes an *Escherichia coli* cell comprising a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ and a mutant TyrRS (LWJ16), wherein the mutant TyrRS (LWJ16) preferentially aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with O-methyl-L-tyrosine in the cell and the cell uses the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ to recognize an amber codon.
- [100] In another preferred embodiment, the invention includes an *Escherichia coli* cell comprising a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ and an SS12-TyrRS, wherein the SS12-TyrRS preferentially aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with L-3-(2-naphthyl)alanine in the cell and the cell uses the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ to recognize an amber codon.
- [101] Sequences of exemplary O-tRNA and O-RS molecules are described in the Examples.

WO 02/085923

PCT/US02/12465

Orthogonal tRNA and Orthogonal aminoacyl-tRNA synthetase pairs

[102] An orthogonal pair is composed of an O-tRNA, e.g., a suppressor tRNA, a frameshift tRNA, or the like, and an O-RS. The O-tRNA is not acylated by endogenous synthetases and is capable of decoding a selector codon, as described above. The O-RS
 5 recognizes the O-tRNA, e.g., with an extended anticodon loop, and preferentially aminoacylates the O-tRNA with an unnatural amino acid. The development of multiple orthogonal tRNA/synthetase pairs can allow the simultaneous incorporation of multiple unnatural amino acids using different codons.

[103] The O-tRNA and the O-RS can be naturally occurring or can be derived by
 10 mutation of a naturally occurring tRNA and/or RS from a variety of organisms, which are described under sources and hosts. In various embodiments, the O-tRNA and O-RS are derived from at least one organism. In another embodiment, the O-tRNA is derived from a naturally occurring or mutated naturally occurring tRNA from a first organism and the O-RS is derived from naturally occurring or mutated naturally occurring RS from a second
 15 organism.

[104] Methods (deriving, mutating, screening) for obtaining O-tRNA, O-RS, and pairs to be used in the compositions and methods of the invention are also described United States Patent application Serial No. _____, titled "Methods and Compositions for the production of orthogonal tRNA-tRNA synthetase pairs," (Attorney docket number 54-
 20 000130US by Schultz et al) filed concurrently with the instant application, the disclosure of which is incorporated in its entirety.

[105] These methods solve the problems discussed in the background section for the other strategies that were attempted to generate orthogonal tRNA/RS pairs. Specifically, these methods include: (a) generating a library of tRNAs derived from at least one tRNA
 25 from a first organism; (b) negatively selecting the library for tRNAs that are aminoacylated by an aminoacyl-tRNA synthetase (RS) from a second organism in the absence of a RS from the first organism, thereby providing a pool of tRNAs; (c) selecting the pool of tRNAs for members that are aminoacylated by an introduced orthogonal RS (O-RS), thereby providing at least one recombinant O-tRNA. The at least one
 30 recombinant O-tRNA recognizes a selector codon and is not efficiently recognized by the RS from the second organism and is preferentially aminoacylated by the O-RS. The method also includes: (d) generating a library of mutant RSs derived from at least one aminoacyl-tRNA synthetase (RS) from a third organism; (e) selecting the library of RSs

WO 02/085923

PCT/US02/12465

for members that preferentially aminoacylate the at least one recombinant O-tRNA in the presence of an unnatural amino acid and a natural amino acid, thereby providing a pool of active RSs; and, (f) negatively selecting the pool for active RSs that preferentially aminoacylate the at least one recombinant O-tRNA in the absence of the unnatural amino acid, thereby providing the at least one specific O-tRNA/O-RS pair, where the at least one specific O-tRNA/O-RS pair comprises at least one recombinant O-RS that is specific for the unnatural amino acid and the at least one recombinant O-tRNA.

[106] One strategy for generating an orthogonal pair involves generating mutant libraries from which to screen and/or select an O-tRNA or O-RS.

10 [107] A second strategy for generating an orthogonal tRNA/synthetase pair involves importing a heterologous tRNA/synthetase pair, e.g., importing a pair from another, e.g., source organism into the host cell. The properties of the heterologous synthetase candidate include, e.g., that it does not charge any host cell tRNA, and the properties of the heterologous tRNA candidate include, e.g., that it is not acylated by any host cell synthetase. In addition, the heterologous tRNA derived from the heterologous tRNA is orthogonal to all host cell synthetases.

[108] Using the methods described herein and in United States Patent application Serial No. _____, titled "Methods and Compositions for the production of orthogonal tRNA-tRNA synthetase pairs," (Attorney docket number 54-000130US, by Schultz et al), 20 the pairs and components of pairs desired above are evolved to generate orthogonal tRNA/synthetase pairs that possess desired characteristic, e.g., that can preferentially aminoacylate an O-tRNA with an unnatural amino acid.

[109] Although discussed with reference to strategies for incorporating unnatural amino acids into proteins in vivo herein, it will be appreciated that strategies can be developed to 25 incorporate natural amino acids in response to selector codons as well, providing an additional basis of and for mutagenesis. That is, a synthetase can be modified to load a natural amino acid onto an orthogonal tRNA that recognizes a selector codon in a manner similar to the loading of an unnatural amino acid as described throughout.

Production of orthogonal aminoacyl tRNA synthetases (O-RS)

30 [110] Methods for producing an O-RS are based on generating a pool of mutant synthetases from the framework of a wild-type synthetase, and then selecting for mutated RSs based on their specificity for an unnatural amino acid relative to the common twenty. To isolate such a synthetase, the selection methods of the present invention are: (i)

WO 02/085923

PCT/US02/12465

sensitive, as the activity of desired synthetases from the initial rounds can be low and the population small; (ii) "tunable", since it is desirable to vary the selection stringency at different selection rounds; and, (iii) general, so that it can be used for different unnatural amino acids.

5 [111] Methods to generate an orthogonal aminoacyl tRNA synthetase include mutating the synthetase, e.g., at the active site in the synthetase, at the editing mechanism site in the synthetase, at different sites by combining different domains of synthetases, or the like, and applying a selection process. A strategy is used, which is based on the combination of a positive selection followed by a negative selection. In the positive selection, suppression
10 of the selector codon introduced at a nonessential position(s) of a positive marker allows cells to survive under positive selection pressure. In the presence of both natural and unnatural amino acids, survivors thus encode active synthetases charging the orthogonal suppressor tRNA with either a natural or unnatural amino acid. In the negative selection, suppression of a selector codon introduced at a nonessential position(s) of a negative
15 marker removes synthetases with natural amino acid specificities. Survivors of the negative and positive selection encode synthetases that aminoacylate (charge) the orthogonal suppressor tRNA with unnatural amino acids only. These synthetases can then be subjected to further mutagenesis, e.g., DNA shuffling or other recursive mutagenesis methods.

20 [112] The library of mutant RSs can be generated using various mutagenesis techniques known in the art. For example, the mutant RSs can be generated by site-specific mutations, random point mutations, homologous recombination, chimeric construction or the like.

[113] The positive selection step can include, e.g., introducing a positive selection
25 marker, e.g., an antibiotic resistance gene, or the like, and the library of mutant RSs into a plurality of cells, wherein the positive selection marker comprises at least one selector codon, e.g., an amber codon; growing the plurality of cells in the presence of a selection agent; selecting cells that survive in the presence of the selection agent by suppressing the at least one selector codon in the positive selection marker, thereby providing a subset of
30 positively selected cells that contains the pool of active mutant RSs. Optionally, the selection agent concentration can be varied.

[114] The negative selection can include, e.g., introducing a negative selection marker with the pool of active mutant RSs from the positive selection into a plurality of cells of a second organism, wherein the negative selection marker is an antibiotic resistance gene,

WO 02/085923

PCT/US02/12465

e.g., a chloramphenicol acetyltransferase (CAT) gene, comprising at least one selector codon; and, selecting cells that survive in a 1st media supplemented with the unnatural amino acid and a selection agent, but fail to survive in a 2nd media not supplemented with the unnatural amino acid and the selection agent, thereby providing surviving cells with the at least one recombinant O-RS. Optionally, the concentration of the selection agent is varied.

[115] The positive selection can be based on suppression of a selector codon in a positive selection marker, e.g., a chloramphenicol acetyltransferase (CAT) gene comprising a selector codon, e.g., an amber stop codon, in the CAT gene, so that chloramphenicol can be applied as the positive selection pressure. In addition, the CAT gene can be used as both a positive marker and negative marker as describe herein in the presence and absence of unnatural amino acid. Optionally, the CAT gene comprising a selector codon is used for the positive selection and a negative selection marker, e.g., a toxic marker, such as a barnase gene comprising at least one or more selector codons, is used for the negative selection.

[116] The positive selection can also be based on suppression of a selector codon at a nonessential position in the β -lactamase gene, rendering cells ampicillin resistant; and a negative selection using the ribonuclease barnase as the negative marker is used. In contrast to β -lactamase, which is secreted into the periplasm, CAT localizes in the cytoplasm; moreover, ampicillin is bacteriocidal, while chloramphenicol is bacteriostatic.

[117] The recombinant O-RS can be further mutated and selected. In one embodiment, the methods for producing at least one recombinant orthogonal aminoacyl-tRNA synthetase (O-RS) can further comprise: (d) isolating the at least one recombinant O-RS; (e) generating a second set of mutated O-RS derived from the at least one recombinant O-RS; and, (f) repeating steps (b) and (c) until a mutated O-RS is obtained that comprises an ability to preferentially aminoacylate the O-tRNA. Optionally, steps (d)-(f) are repeated, e.g., at least about two times. In one aspect, the second set of mutated O-RS can be generated by mutagenesis, e.g., random mutagenesis, site-specific mutagenesis, recombination or a combination thereof.

30 Production of orthogonal tRNA (O-tRNAs)

[118] Methods for producing a recombinant orthogonal tRNA (O-tRNA) are provided in United States Patent application Serial No. _____, titled "Methods and Compositions

WO 02/085923

PCT/US02/12465

for the production of orthogonal tRNA-tRNA synthetase pairs," (Attorney docket number 54-000130US, by Schultz et al).

- [119] Methods of producing a recombinant O-tRNA include: (a) generating a library of mutant tRNAs derived from at least one tRNA, e.g., a suppressor tRNA, from a first organism; (b) negatively selecting the library for mutant tRNAs that are aminoacylated by an aminoacyl-tRNA synthetase (RS) from a second organism in the absence of a RS from the first organism, thereby providing a pool of mutant tRNAs; and, (c) selecting the pool of mutant tRNAs for members that are aminoacylated by an introduced orthogonal RS (O-RS), thereby providing at least one recombinant O-tRNA; wherein the at least one recombinant O-tRNA recognizes a selector codon and is not efficiently recognized by the RS from the second organism and is preferentially aminoacylated by the O-RS. In one embodiment, the recombinant O-tRNA possesses an improvement of orthogonality.
- [120] For example, to improve the orthogonality of a tRNA while preserving its affinity toward a desired RS, the methods include a combination of negative and positive selections with a mutant suppressor tRNA library in the absence and presence of the cognate synthetase, respectively. In the negative selection, a selector codon(s) is introduced in a marker gene, e.g., a toxic gene, such as barnase, at a nonessential position. When a member of the mutated tRNA library, e.g., derived from *Methanococcus jannaschii*, is aminoacylated by endogenous host, e.g., *Escherichia coli* synthetases (i.e., it is not orthogonal to the host, e.g., *Escherichia coli* synthetases), the selector codon, e.g., an amber codon, is suppressed and the toxic gene product produced leads to cell death. Cells harboring orthogonal tRNAs or non-functional tRNAs survive. Survivors are then subjected to a positive selection in which a selector codon, e.g., an amber codon, is placed in a positive marker gene, e.g., a drug resistance gene, such as a β -lactamase gene. These cells also contain an expression vector with a cognate RS. These cells are grown in the presence of a selection agent, e.g., ampicillin. tRNAs are then selected for their ability to be aminoacylated by the coexpressed cognate synthetase and to insert an amino acid in response to this selector codon. Cells harboring non-functional tRNAs, or tRNAs that cannot be recognized by the synthetase of interest are sensitive to the antibiotic.
- Therefore, tRNAs that: (i) are not substrates for endogenous host, e.g., *Escherichia coli*, synthetases; (ii) can be aminoacylated by the synthetase of interest; and (iii) are functional in translation survive both selections.

WO 02/085923

PCT/US02/12465

[121] Libraries of mutated tRNA are constructed. Mutations can be introduced at a specific position(s), e.g., at a nonconservative position(s), or at a conservative position, at a randomized position(s), or a combination of both in a desired loop of a tRNA, e.g., an anticodon loop, (D arm, V loop, T Ψ C arm) or a combination of loops or all loops.

- 5 Chimeric libraries of tRNA are also included in the present invention. It should be noted that libraries of tRNA synthetases from various organism (e.g., microorganisms such as eubacteria or archaeobacteria) such as libraries that comprise natural diversity (see, e.g., U.S. Patent No. 6,238,884 to Short et al; U.S. Patent No. 5,756,316 to Schallenberger et al; U.S. Patent No. 5,783,431 to Petersen et al; U.S. Patent No. 5,824,485 to Thompson et al; 10 U.S. Patent No. 5,958,672 to Short et al), are optionally constructed and screened for orthogonal pairs.

- [122] For example, negatively selecting the library for mutant tRNAs that are aminoacylated by an aminoacyl-tRNA synthetase can include: introducing a toxic marker gene, wherein the toxic marker gene comprises at least one of the selector codons and the 15 library of mutant tRNAs into a plurality of cells from the second organism; and, selecting surviving cells, wherein the surviving cells contain the pool of mutant tRNAs comprising at least one orthogonal tRNA or nonfunctional tRNA. For example, the toxic marker gene is a ribonuclease barnase gene, wherein the ribonuclease barnase gene comprises at least one amber codon. Optionally, the ribonuclease barnase gene can include two or more 20 amber codons. The surviving cells can be selected, e.g., by using a comparison ratio cell density assay.

- [123] In another example, selecting the pool of mutant tRNAs for members that are aminoacylated by an introduced orthogonal RS (O-RS) can include: introducing a positive selection marker gene, wherein the positive selection marker gene comprises a drug 25 resistance gene, e.g., a β -lactamase gene, comprising at least one of the selector codons, e.g., a β -lactamase gene comprising at least one amber stop codon, the O-RS, and the pool of mutant tRNAs into a plurality of cells from the second organism; and, selecting surviving cells grown in the presence of a selection agent, e.g., an antibiotic, thereby providing a pool of cells possessing the at least one recombinant tRNA, wherein the 30 recombinant tRNA is aminoacylated by the O-RS and inserts an amino acid into a translation product encoded by the positive marker gene, in response to the at least one selector codons. In another embodiment, the concentration of the selection agent is varied. Recombinant O-tRNAs produced by the methods are included in the present invention.

WO 02/085923

PCT/US02/12465

- [124] The stringency of the selection steps, e.g., the positive selection step, the negative selection step or both the positive and negative selection steps, in the above described methods, optionally include varying the selection stringency. For example, because barnase is an extremely toxic protein, the stringency of the negative selection can be controlled by introducing different numbers of selector codons into the barnase gene. In one aspect of the present invention, the stringency is varied because the desired activity can be low during early rounds. Thus, less stringent selection criteria are applied in early rounds and more stringent criteria are applied in later rounds of selection.
- [125] Other types of selections can be used in the present invention for generating, e.g., O-RS, O-tRNA, and O-tRNA/O-RS pairs. For example, the positive selection step, the negative selection step or both the positive and negative selection steps can include using a reporter, wherein the reporter is detected by fluorescence-activated cell sorting (FACS). For example, a positive selection can be done first with a positive selection marker, e.g., chloramphenicol acetyltransferase (CAT) gene, where the CAT gene comprises a selector codon, e.g., an amber stop codon, in the CAT gene, which followed by a negative selection screen, that is based on the inability to suppress a selector codon(s), e.g., two or more, at positions within a negative marker, e.g., T7 RNA polymerase gene. In one embodiment, the positive selection marker and the negative selection marker can be found on the same vector, e.g., plasmid. Expression of the negative marker drives expression of the reporter, e.g., green fluorescent protein (GFP). The stringency of the selection and screen can be varied, e.g., the intensity of the light need to fluorescence the reporter can be varied. In another embodiment, a positive selection can be done with a reporter as a positive selection marker, which is screened by FACS, followed by a negative selection screen, that is based on the inability to suppress a selector codon(s), e.g., two or more, at positions within a negative marker, e.g., barnase gene.
- [126] Optionally, the reporter is displayed on a cell surface, on a phage display or the like. Cell-surface display, e.g., the OmpA-based cell-surface display system, relies on the expression of a particular epitope, e.g., a poliovirus C3 peptide fused to an outer membrane porin OmpA, on the surface of the *Escherichia coli* cell. The epitope is displayed on the cell surface only when a selector codon in the protein message is suppressed during translation. The displayed peptide then contains the amino acid recognized by one of the mutant aminoacyl-tRNA synthetases in the library, and the cell containing the corresponding synthetase gene can be isolated with antibodies raised against peptides containing specific unnatural amino acids. The OmpA-based cell-surface

WO 02/085923

PCT/US02/12465

display system was developed and optimized by Georgiou et al. as an alternative to phage display. See, Francisco, J. A., Campbell, R., Iverson, B. L. & Georgiou, G. *Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface*. *Proc Natl Acad Sci U S A*. 90:10444-8 (1993).

- 5 [127] The selection steps can also be carried out in vitro. The selected component, e.g., synthetase and/or tRNA, can then be introduced into a cell for use in in vivo incorporation of an unnatural amino acid.

Source and Host Organisms

- [128] The orthogonal tRNA-RS pair, e.g., derived from at least a first, e.g., source
10 organism or at least two source organisms, which can be the same or different, can be used in a variety of host organisms, e.g., a second organism. The first and the second organisms of the methods of the present invention can be the same or different. In one embodiment, the first organism is a prokaryotic organism, e.g., *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, *Escherichia coli*, *A. fulgidus*,
15 *Halobacterium*, *P. furiosus*, *P. horikoshii*, *A. pernix*, *T. thermophilus*, or the like. Alternatively, the first organism is a eukaryotic organism, e.g., plants (e.g., complex plants such as monocots, or dicots), algae, protists, fungi (e.g., yeast, etc.), animals (e.g., mammals, insects, arthropods, etc.), or the like. In another embodiment, the second organism is a prokaryotic organism, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, *Escherichia coli*, *A. fulgidus*, *Halobacterium*, *P.*
20 *furiosus*, *P. horikoshii*, *A. pernix*, *T. thermophilus*, or the like. Alternatively, the second organism can be a eukaryotic organism, e.g., plants, fungi, animals, or the like.
- [129] As described above, the individual components of a pair can be derived from the same organism or different organisms. For example, tRNA can be derived from a
25 prokaryotic organism, e.g., an archaeobacterium, such as *Methanococcus jannaschii* and *Halobacterium NRC-1* or a eubacterium, such as *Escherichia coli*, while the synthetase can be derived from same or another prokaryotic organism, such as, *Methanococcus jannaschii*, *Archaeoglobus fulgidus*, *Methanobacterium thermoautotrophicum*, *P. furiosus*, *P. horikoshii*, *A. pernix*, *T. thermophilus*, *Halobacterium*, *Escherichia coli* or the like.
- 30 Eukaryotic sources can also be used, e.g., plants (e.g., complex plants such as monocots, or dicots), algae, protists, fungi (e.g., yeast, etc.), animals (e.g., mammals, insects, arthropods, etc.), or the like.

WO 02/085923

PCT/US02/12465

Selector Codons

[130] Selector codons of the present invention expand the genetic codon framework of protein biosynthetic machinery. For example, a selector codon includes, e.g., a unique three base codon, a nonsense codon, such as a stop codon, e.g., an amber codon, or an opal
5 codon, an unnatural codon, at least a four base codon or the like. A number of selector codons can be introduced into a desired gene, e.g., one or more, two or more, more than three, etc.

[131] The 64 genetic codons code for 20 amino acids and 3 stop codons. Because only one stop codon is needed for translational termination, the other two can in principle be
10 used to encode nonproteinogenic amino acids. The amber stop codon, UAG, has been successfully used in in vitro biosynthetic system and in *Xenopus* oocytes to direct the incorporation of unnatural amino acids. Among the 3 stop codons, UAG is the least used stop codon in *Escherichia coli*. Some *Escherichia coli* strains contain natural suppressor tRNAs, which recognize UAG and insert a natural amino acid. In addition, these amber
15 suppressor tRNAs have been used in conventional protein mutagenesis.

[132] In one embodiment, the methods involve the use of a selector codon that is a stop codon for the incorporation of unnatural amino acids in vivo. For example, an O-tRNA is generated that recognizes the stop codon, e.g., UAG, and is aminoacylated by an O-RS with a desired unnatural amino acid. This O-tRNA is not recognized by the naturally
20 occurring aminoacyl-tRNA synthetases. Conventional site-directed mutagenesis can be used to introduce the stop codon, e.g., TAG, at the site of interest in the protein gene. See, e.g., Sayers, J.R., Schmidt, W., Eckstein, F. 5',3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res*, 791-802 (1988). When the O-RS, O-tRNA and the mutant gene are combined in vivo, the unnatural amino acid is
25 incorporated in response to the UAG codon to give a protein containing the unnatural amino acid at the specified position.

[133] The incorporation of unnatural amino acids in vivo can be done without significant perturbation of the host, e.g., *Escherichia coli*. For example, because the suppression efficiency for the UAG codon depends upon the competition between the O-tRNA, e.g.,
30 the amber suppressor tRNA, and the release factor 1 (RF1) (which binds to the UAG codon and initiates release of the growing peptide from the ribosome), the suppression efficiency can be modulated by, e.g., either increasing the expression level of O-tRNA, e.g., the suppressor tRNA, or using an RF1 deficient strain.

WO 02/085923

PCT/US02/12465

- [134] Unnatural amino acids can also be encoded with rare codons. For example, when the arginine concentration in an in vitro protein synthesis reaction is reduced, the rare arginine codon, AGG, has proven to be efficient for insertion of Ala by a synthetic tRNA acylated with alanine. See, e.g., C. H. Ma, W. Kudlicki, O. W. Odom, G. Kramer and B. Hardesty, *Biochemistry*, 32:7939 (1993). In this case, the synthetic tRNA competes with the naturally occurring tRNA^{Arg}, which exists as a minor species in *Escherichia coli*. Some organisms do not use all triplet codons. An unassigned codon AGA in *Micrococcus luteus* has been utilized for insertion of amino acids in an in vitro transcription/translation extract. See, e.g., A. K. Kowal and J. S. Oliver, *Nucl. Acid. Res.*, 25:4685 (1997).
- 10 Components of the present invention can be generated to use these rare codons in vivo.
- [135] Selector codons also comprise four or more base codons, such as, four, five, six or more base codons. Examples of four base codons include, e.g., AGGA, CUAG, UAGA, CCCU and the like. Examples of five base codons include, e.g., AGGAC, CCCCU, CCCUC, CUAGA, CUACU, UAGGC and the like. For example, in the presence of
- 15 mutated O-tRNAs, e.g., a special frameshift suppressor tRNAs, with anticodon loops, e.g., with at least 8-10 nt anticodon loops, the four or more base codon is read as single amino acid. In other embodiments, the anticodon loops can decode, e.g., at least a four-base codon, at least a five-base codon, or at least a six-base codon or more. Since there are 256 possible four-base codons, multiple unnatural amino acids can be encoded in the same cell
- 20 using the four or more base codon. See, J. Christopher Anderson et al., Exploring the Limits of Codon and Anticodon Size, *Chemistry and Biology*, Vol. 9, 237-244 (2002); Thomas J. Magliery, Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in *Escherichia coli*, *J. Mol. Biol.* 307: 755-769 (2001).
- 25 [136] Methods of the present invention include using extended codons based on frameshift suppression. Four or more base codons can insert, e.g., one or multiple unnatural amino acids into the same protein. For example, four-base codons have been used to incorporate unnatural amino acids into proteins using in vitro biosynthetic methods. See, e.g., C. H. Ma, W. Kudlicki, O. W. Odom, G. Kramer and B. Hardesty,
- 30 *Biochemistry*, 1993, 32, 7939 (1993); and, T. Hoshaka, D. Kajihara, Y. Ashizuka, H. Murakami and M. Sisido, *J. Am. Chem. Soc.*, 121:34 (1999). CGGG and AGGU were used to simultaneously incorporate 2-naphthylalanine and an NBD derivative of lysine into streptavidin in vitro with two chemically acylated frameshift suppressor tRNAs. See, e.g., T. Hoshaka, Y. Ashizuka, H. Sasaki, H. Murakami and M. Sisido, *J. Am. Chem. Soc.*,

WO 02/085923

PCT/US02/12465

121:12194 (1999). In an in vivo study, Moore et al. examined the ability of tRNA^{Leu} derivatives with NCUA anticodons to suppress UAGN codons (N can be U, A, G, or C), and found that the quadruplet UAGA can be decoded by a tRNA^{Leu} with a UCUA anticodon with an efficiency of 13 to 26% with little decoding in the 0 or -1 frame. See,

- 5 B. Moore, B. C. Persson, C. C. Nelson, R. F. Gesteland and J. F. Atkins, *J. Mol. Biol.*, 298:195 (2000). In one embodiment, extended codons based on rare codons or nonsense codons can be used in present invention, which can reduce missense readthrough and frameshift suppression at other unwanted sites.

[137] A translational bypassing system can also be used to incorporate an unnatural amino acid in a desired polypeptide. In a translational bypassing system, a large sequence is inserted into a gene but is not translated into protein. The sequence contains a structure that serves as a cue to induce the ribosome to hop over the sequence and resume translation downstream of the insertion.

- [138] Alternatively, or in combination with others methods described above to
15 incorporate an unnatural amino acid in a polypeptide, a trans-translation system can be used. This system involves a molecule called tmRNA present in *Escherichia coli*. This RNA molecule is structurally related to an alanyl tRNA and is aminoacylated by the alanyl synthetase. The difference between tmRNA and tRNA is that the anticodon loop is replaced with a special large sequence. This sequence allows the ribosome to resume
20 translation on sequences that have stalled using an open reading frame encoded within the tmRNA as template. In the present invention, an orthogonal tmRNA can be generated that is preferentially aminoacylated with an orthogonal synthetase and loaded with an unnatural amino acid. By transcribing a gene using the system, the ribosome stalls at a specific site; the unnatural amino acid is introduced at that site, then translation resumes,
25 using the sequence encoded within the orthogonal tmRNA.

- [139] Selector codons optionally include unnatural base pairs. These unnatural base pairs further expand the existing genetic alphabet. One extra base pair increases the number of triplet codons from 64 to 125. Properties of third base pairs include stable and selective base pairing, efficient enzymatic incorporation into DNA with high fidelity by a
30 polymerase, and the efficient continued primer extension after synthesis of the nascent unnatural base pair. Descriptions of unnatural base pairs which can be adapted for methods and compositions include, e.g., Hirao, et al., An unnatural base pair for incorporating amino acid analogues into protein, *Nature Biotechnology*, 20:177-182 (2002). Other relevant publications are listed below.

WO 02/085923

PCT/US02/12465

[140] For in vivo usage, the unnatural nucleoside is membrane permeable and is phosphorylated to form the corresponding triphosphate. In addition, the increased genetic information is stable and not destroyed by cellular enzymes. Previous efforts by Benner and others took advantage of hydrogen bonding patterns that are different from those in canonical Watson-Crick pairs, the most noteworthy example of which is the iso-C:iso-G pair. See, e.g., C. Switzer, S. E. Moroney and S. A. Benner, *J. Am. Chem. Soc.*, 111:8322 (1989); and, J. A. Piccirilli, T. Krauch, S. E. Moroney and S. A. Benner, *Nature*, 1990, 343:33 (1990); E. T. Kool, *Curr. Opin. Chem. Biol.*, 4:602 (2000). These bases in general mispair to some degree with natural bases and cannot be enzymatically replicated. Kool and co-workers demonstrated that hydrophobic packing interactions between bases can replace hydrogen bonding to drive the formation of base pair. See, E. T. Kool, *Curr. Opin. Chem. Biol.*, 4:602 (2000); and, K. M. Guckian and E. T. Kool, *Angew. Chem. Int. Ed. Engl.*, 36, 2825 (1998). In an effort to develop an unnatural base pair satisfying all the above requirements, Schultz, Romesberg and co-workers have systematically synthesized and studied a series of unnatural hydrophobic bases. A PICS:PICS self-pair is found to be more stable than natural base pairs, and can be efficiently incorporated into DNA by Klenow fragment of *Escherichia coli* DNA polymerase I (KF). See, e.g., D. L. McMinn, A. K. Ogawa, Y. Q. Wu, J. Q. Liu, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 121:11586 (1999); and, A. K. Ogawa, Y. Q. Wu, D. L. McMinn, J. Q. Liu, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 122:3274 (2000). A 3MN:3MN self-pair can be synthesized by KF with efficiency and selectivity sufficient for biological function. See, e.g., A. K. Ogawa, Y. Q. Wu, M. Berger, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 122:8803 (2000). However, both bases act as a chain terminator for further replication. A mutant DNA polymerase has been recently evolved that can be used to replicate the PICS self pair. In addition, a 7AI self pair can be replicated. See, e.g., E. J. L. Tae, Y. Q. Wu, G. Xia, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 123:7439 (2001). A novel metallobase pair, Dipic:Py, has also been developed, which forms a stable pair upon binding Cu(II). See, E. Meggers, P. L. Holland, W. B. Tolman, F. E. Romesberg and P. G. Schultz, *J. Am. Chem. Soc.*, 122:10714 (2000). Because extended codons and unnatural codons are intrinsically orthogonal to natural codons, the methods of the present invention can take advantage of this property to generate orthogonal tRNAs for them.

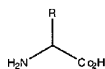
WO 02/085923

PCT/US02/12465

Unnatural amino acids

[141] As used herein an unnatural amino acid refers to any amino acid, modified amino acid, or amino acid analogue other than selenocysteine and the following twenty genetically encoded alpha-amino acids: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine. The generic structure of an alpha-amino acid is illustrated by Formula I:

I



[142]

[143] An unnatural amino acid is typically any structure having Formula I wherein the R group is any substituent other than one used in the twenty natural amino acids. See, e.g., *Biochemistry* by L. Stryer, 3rd ed. 1988, Freeman and Company, New York, for structures of the twenty natural amino acids. Note that, the unnatural amino acids of the present invention can be naturally occurring compounds other than the twenty alpha-amino acids above.

[144] Because the unnatural amino acids of the invention typically differ from the natural amino acids in side chain only, the unnatural amino acids form amide bonds with other amino acids, e.g., natural or unnatural, in the same manner in which they are formed in naturally occurring proteins. However, the unnatural amino acids have side chain groups that distinguish them from the natural amino acids. For example, R in Formula I optionally comprises an alkyl-, aryl-, acyl-, keto-, azido-, hydroxyl-, hydrazine, cyano-, halo-, hydrazide, alkenyl, alkynyl, ether, thiol, seleno-, sulfonyl-, borate, boronate, phospho-, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, ester, thioacid, hydroxylamine, amino group, or the like or any combination thereof. Other unnatural amino acids of interest include, but are not limited to, amino acids comprising a photoactivatable cross-linker, spin-labeled amino acids, fluorescent amino acids, metal binding amino acids, metal-containing amino acids, radioactive amino acids, amino acids with novel functional groups, amino acids that covalently or noncovalently interact with other molecules, photocaged and/or photoisomerizable amino acids, amino acids comprising biotin or a biotin analogue, glycosylated amino acids such as a sugar substituted serine, other carbohydrate modified amino acids, keto containing amino acids,

WO 02/085923

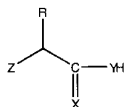
PCT/US02/12465

amino acids comprising polyethylene glycol or polyether, heavy atom substituted amino acids, chemically cleavable and/or photocleavable amino acids, amino acids with an elongated side chains as compared to natural amino acids, e.g., polyethers or long chain hydrocarbons, e.g., greater than about 5 or greater than about 10 carbons, carbon-linked sugar-containing amino acids, redox-active amino acids, amino thioacid containing amino acids, and amino acids comprising one or more toxic moiety.

[145] In addition to unnatural amino acids that contain novel side chains, unnatural amino acids also optionally comprise modified backbone structures, e.g., as illustrated by the structures of Formula II and III:

10 [146]

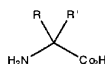
II



[147]

[148]

III



[149]

[150] wherein Z typically comprises OH, NH₂, SH, NH-R', or S-R'; X and Y, which can be the same or different, typically comprise S or O, and R and R', which are optionally the same or different, are typically selected from the same list of constituents for the R group described above for the unnatural amino acids having Formula I as well as hydrogen. For example, unnatural amino acids of the invention optionally comprise substitutions in the amino or carboxyl group as illustrated by Formulas II and III. Unnatural amino acids of this type include, but are not limited to, α-hydroxy acids, α-thioacids α-aminothiocarboxylates, e.g., with side chains corresponding to the common twenty natural amino acids or unnatural side chains. In addition, substitutions at the α-carbon optionally include L, D, or α-α-disubstituted amino acids such as D-glutamate, D-alanine, D-methyl-O-tyrosine, aminobutyric acid, and the like. Other structural alternatives include cyclic amino acids, such as proline analogues as well as 3,4,6,7,8, and 9 membered ring proline analogues, β and γ amino acids such as substituted β-alanine and γ-amino butyric acid.

WO 02/085923

PCT/US02/12465

[151] For example, many unnatural amino acids are based on natural amino acids, such as tyrosine, glutamine, phenylalanine, and the like. Tyrosine analogs include para-substituted tyrosines, ortho-substituted tyrosines, and meta substituted tyrosines, wherein the substituted tyrosine comprises an acetyl group, a benzoyl group, an amino group, a hydrazine, an hydroxyamine, a thiol group, a carboxy group, an isopropyl group, a methyl group, a C₆ - C₂₀ straight chain or branched hydrocarbon, a saturated or unsaturated hydrocarbon, an O-methyl group, a polyether group, a nitro group, or the like. In addition, multiply substituted aryl rings are also contemplated. Glutamine analogs of the invention include, but are not limited to, α -hydroxy derivatives, γ -substituted derivatives, cyclic derivatives, and amide substituted glutamine derivatives. Example phenylalanine analogs include, but are not limited to, meta-substituted phenylalanines, wherein the substituent comprises a hydroxy group, a methoxy group, a methyl group, an allyl group, an acetyl group, or the like. Specific examples of unnatural amino acids include, but are not limited to, O-methyl-L-tyrosine, an L-3-(2-naphthyl)alanine, a 3-methyl-phenylalanine, an O-4-allyl-L-tyrosine, a 4-propyl-L-tyrosine, a tri-O-acetyl-GlcNAc β -serine, an L-Dopa, a fluorinated phenylalanine, an isopropyl-L-phenylalanine, a *p*-azido-L-phenylalanine, a *p*-acyl-L-phenylalanine, a *p*-benzoyl-L-phenylalanine, an L-phosphoserine, a phosphoserine, a phosphotyrosine, a *p*-iodo-phenylalanine, a *p*-bromophenylalanine, a *p*-amino-L-phenylalanine, and an isopropyl-L-phenylalanine, and the like. The structures of a variety of unnatural amino acids are provided in the figures, e.g., Figures 17, 18, 19, 26, and 29.

[152] Typically, the unnatural amino acids of the invention are selected or designed to provide additional characteristics unavailable in the twenty natural amino acids. For example, unnatural amino acid are optionally designed or selected to modify the biological properties of a protein, e.g., into which they are incorporated. For example, the following properties are optionally modified by inclusion of an unnatural amino acid into a protein: toxicity, biodistribution, solubility, stability, e.g., thermal, hydrolytic, oxidative, resistance to enzymatic degradation, and the like, facility of purification and processing, structural properties, spectroscopic properties, chemical and/or photochemical properties, catalytic activity, redox potential, half-life, ability to react with other molecules, e.g., covalently or noncovalently, and the like.

WO 02/085923

PCT/US02/12465

Chemical Synthesis of Unnatural Amino Acids

- [153] Many of the unnatural amino acids provided above are commercially available, e.g., from Sigma (USA) or Aldrich (Milwaukee, WI, USA). Those that are not commercially available are optionally synthesized as provided in the examples below or using standard methods known to those of skill in the art. For organic synthesis techniques, see, e.g., Organic Chemistry by Fessenden and Fessenden, (1982, Second Edition, Willard Grant Press, Boston Mass.); Advanced Organic Chemistry by March (Third Edition, 1985, Wiley and Sons, New York); and Advanced Organic Chemistry by Carey and Sundberg (Third Edition, Parts A and B, 1990, Plenum Press, New York).
- 5
- 10 [154] For example, meta-substituted phenylalanines are synthesized in a procedure as outlined in Figure 14. Typically, NBS (N-bromosuccinimide) is added to a meta-substituted methylbenzene compound to give a meta-substituted benzyl bromide, which is then reacted with a malonate compound to give the meta substituted phenylalanine. Typical substituents used for the meta position include, but are not limited to, ketones, methoxy groups, alkyls, acetyls, and the like. For example, 3-acetyl-phenylalanine is made by reacting NBS with a solution of 3-methylacetophenone. For more details see the examples below. A similar synthesis is used to produce a 3-methoxy phenylalanine. The R group on the meta position of the benzyl bromide in that case is -OCH₃. See, e.g., Matsoukas et al., *J. Med. Chem.*, 1995, 38, 4660-4669.
- 15
- 20 [155] In some embodiments, the design of unnatural amino acids is biased by known information about the active sites of synthetases, e.g., orthogonal tRNA synthetases used to aminoacylate an orthogonal tRNA. For example, three classes of glutamine analogs are provided, including derivatives substituted at the nitrogen of amide (1), a methyl group at the γ -position (2), and a N-C γ -cyclic derivative (3). Based upon the x-ray crystal structure of *E. coli* GlnRS, in which the key binding site residues are homologous to yeast GlnRS, the analogs were designed to complement an array of side chain mutations of residues within a 10 Å shell of the side chain of glutamine, e.g., a mutation of the active site Phe233 to a small hydrophobic amino acid might be complemented by increased steric bulk at the C γ position of Gln.
- 25
- 30 [156] For example, N-phthaloyl-L-glutamic 1,5-anhydride (compound number 4 in Figure 23) is optionally used to synthesize glutamine analogs with substituents at the nitrogen of the amide. See, e.g., King, F.E. & Kidd, D.A.A. A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates. *J. Chem.*

WO 02/085923

PCT/US02/12465

- Soc., 3315-3319 (1949); Friedman, O.M. & Chatterji, R. Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents. *J. Am. Chem. Soc.* 81, 3750-3752 (1959); Craig, J.C. et al. Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine). *J. Org. Chem.* 53, 1167-1170 (1988); and Azoulay, M., Vilmon, M. & Frappier, F. Glutamine analogues as Potential Antimalarials, *Eur. J. Med. Chem.* 26, 201-5 (1991). The anhydride is typically prepared from glutamic acid by first protection of the amine as the phthalimide followed by refluxing in acetic acid. The anhydride is then opened with a number of amines, resulting in a range of substituents at the amide. Deprotection of the phthaloyl group with hydrazine affords a free amino acid as shown in Figure 23.
- [157] Substitution at the γ -position is typically accomplished via alkylation of glutamic acid. See, e.g., Koskinen, A.M.P. & Rapoport, H. Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues. *J. Org. Chem.* 54, 1859-1866. (1989). A protected amino acid, e.g., as illustrated by compound number 5 in Figure 24 is optionally prepared by first alkylation of the amino moiety with 9-bromo-9-phenylfluorene (PhflBr) (see, e.g., Christie, B.D. & Rapoport, H. Synthesis of Optically Pure Pivcolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization. *J. Org. Chem.* 1989, 1859-1866 (1985)) and then esterification of the acid moiety using O-tert-butyl-N, N'-diisopropylisourea. Addition of $\text{KN}(\text{Si}(\text{CH}_3)_3)_2$ regioselectively deprotonates at the α -position of the methyl ester to form the enolate, which is then optionally alkylated with a range of alkyl iodides. Hydrolysis of the t-butyl ester and Phfl group gave the desired γ -methyl glutamine analog (Compound number 2 in Figure 24).
- [158] An N-C' cyclic analog, as illustrated by Compound number 3 in Figure 25, is optionally prepared in 4 steps from Boc-Asp-Ot-Bu as previously described. See, e.g., Barton, D.H.R., Herve, Y., Potier, P. & Thierry, J. Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- α -Amino-Adipic Acids, L- α -aminopimelic Acid and Appropriate Unsaturated Derivatives. *Tetrahedron Lett.* 43, 4297-4308 (1987) and, Subasinghe, N., Schulte, M., Roon, R.J., Koerner, J.F. & Johnson, R.L. Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site. *J. Med. Chem.* 35 4602-7 (1992). Generation of the anion of the N-t-Boc-pyrrolidinone, pyrrolidinone, or

WO 02/085923

PCT/US02/12465

oxazolidone followed by the addition of the compound 7, as shown in Figure 25, results in a Michael addition product. Deprotection with TFA then results in the free amino acids.

[159] In addition to the above unnatural amino acids, a library of tyrosine analogs has also been designed. Based upon the crystal structure of *B. stearothermophilus* TyrRS, whose active site is highly homologous to that of the *M. jannashii* synthetase, residues within a 10Å shell of the aromatic side chain of tyrosine were mutated (Y32, G34, L65, Q155, D158, A167, Y32 and D158). The library of tyrosine analogs, as shown in Figure 26, has been designed to complement an array of substitutions to these active site amino acids. These include a variety of phenyl substitution patterns, which offer different hydrophobic and hydrogen-bonding properties. Tyrosine analogs are optionally prepared using the general strategy illustrated by Figure 27. For example, an enolate of diethyl acetamidomalonnate is optionally generated using sodium ethoxide. A desired tyrosine analog can then be prepared by adding an appropriate benzyl bromide followed by hydrolysis.

15 Cellular uptake of unnatural amino acids

[160] Unnatural amino acid uptake is one issue that is typically considered when designing and selecting unnatural amino acids, e.g., for incorporation into a protein. For example, the high charge density of α -amino acids suggests that these compounds are unlikely to be cell permeable. Natural amino acids are taken up into bacteria via a collection of protein-based transport systems displaying varying degrees of amino acid specificity. The present invention therefore provides a rapid screen for assessing which unnatural amino acids, if any, are taken up by cells.

[161] For example, a variety of unnatural amino acids are optionally screened in minimal media for toxicity to cells. Toxicities are typically sorted into five groups: (1) no toxicity, in which no significant change in doubling times occurs; (2) low toxicity, in which doubling times increase by less than about 10%; (3) moderate toxicity, in which doubling times increase by about 10% to about 50%; (4) high toxicity, in which doubling times increase by about 50% to about 100%; and (5) extreme toxicity, in which doubling times increase by more than about 100%. See, e.g., Liu, D.R. & Schultz, P.G. Progress toward the evolution of an organism with an expanded genetic code. *Proceedings of the National Academy of Sciences of the United States of America* 96, 4780-4785 (1999). The toxicity of the amino acids scoring as highly or extremely toxic is typically measured as a function of their concentration to obtain IC50 values. In general, amino acids which are very close

WO 02/085923

PCT/US02/12465

analogs of natural amino acids or which display reactive functionality demonstrate the highest toxicities. The former trend suggests that mechanisms of toxicity for these unnatural amino acids can be incorporation into proteins or inhibition of essential enzymes that process natural amino acids.

- 5 [162] To identify possible uptake pathways for toxic amino acids, toxicity assays are optionally repeated at IC50 levels, e.g., in media supplemented with an excess of a structurally similar natural amino acid. For toxic amino acids, the presence of excess natural amino acid typically rescues the ability of the cells to grow in the presence of the toxin, presumably because the natural amino acid effectively outcompetes the toxin for
10 either cellular uptake or for binding to essential enzymes. In these cases, the toxic amino acid is optionally assigned a possible uptake pathway and labeled a "lethal allele" whose complementation is required for cell survival. These lethal alleles are extremely useful for assaying the ability of cells to uptake nontoxic unnatural amino acids. Complementation of the toxic allele, evidenced by the restoration of cell growth, suggests that the nontoxic
15 amino acid is taken up by the cell, possibly by the same uptake pathway as that assigned to the lethal allele. A lack of complementation is inconclusive. For example studies and conclusions see the examples provided below.

- [163] Results obtained, e.g., as described in the examples below, demonstrate that complementation of lethal unnatural amino acid alleles is an efficient method for
20 qualitatively assessing amino acid uptake. The method typically requires far less effort than radiolabeling large numbers of compounds and is therefore a more advantageous method for analyzing unnatural; amino acids of interest. This general strategy is optionally used to rapidly evaluate the cellular uptake of a wide range of molecules such as nucleic acid base analogs, carbohydrate analogs, or peptide analogs. For example, this
25 strategy is optionally used to evaluate the cellular uptake of the unnatural amino acids presented herein.

- [164] The present invention also provides a general method for delivering unnatural amino acids, which is independent of all amino acid uptake pathways. This general method relies on uptake via peptide permeases, which transport dipeptides and tripeptides
30 across the cytoplasmic membrane. Peptide permeases are not very side-chain specific, and the KD values for their substrates are comparable to KD values of amino acid permeases, e.g., about 0.1 mM to about 10 mM). See, e.g., Nickitenko, A., Trakhanov, S. & Quioco, S. A structure of DppA, a periplasmic depeptide transport/chemosensory receptor.

Biochemistry 34, 16585-16595 (1995) and Dunten, P., Mowbray, S.L. Crystal structure of

WO 02/085923

PCT/US02/12465

the dipeptide binding protein from *Escherichia coli* involved in active transport and chemotaxis. *Protein Science* 4, 2327-34 (1995). The unnatural amino acids are then taken up as conjugates of natural amino acids, such as lysine, and released into the cytoplasm upon hydrolysis of the dipeptide by one of endogenous *E. coli* peptidases. To test this approach, we synthesized several Unn-Lys and Lys-Unn dipeptides by solid phase synthesis, and tested the growth of an *E. coli* strain deficient in lysine biosynthesis on lysine minimal media in the presence and absence of these dipeptides. The only source of lysine available to these cells is the dipeptide containing the unnatural amino acid. Uptake of phosphoserine, phosphotyrosine, pentafluorophenylalanine, and caged serine have been analyzed in this manner. In all four cases, growth was observed on 10 mM and higher dipeptide concentrations. Although uptake is easily analyzed with the method provided herein, an alternative to designing unnatural amino acid that are amenable to cellular uptake pathways, is to provide biosynthetic pathways to create amino acids *in vivo*.

15 Biosynthesis of Unnatural Amino Acids

[165] Many biosynthetic pathways already exist in cells for the production of amino acids and other compounds. While a biosynthetic method for a particular unnatural amino acid may not exist in nature, e.g., in *E. coli*, the present invention provides such methods. For example, biosynthetic pathways for unnatural amino acids are optionally generated in *E. coli* by adding new enzymes or modifying existing *E. coli* pathways. Additional new enzymes are optionally naturally occurring enzymes or artificially evolved enzymes. For example, the biosynthesis of *p*-aminophenylalanine (as presented in an example below) relies on the addition of a combination of known enzymes from other organisms. The genes for these enzymes can be introduced into a cell, e.g., an *E. coli* cell, by transforming the cell with a plasmid comprising the genes. The genes, when expressed in the cell, provide an enzymatic pathway to synthesize the desired compound. Examples of the types of enzymes that are optionally added are provided in the examples below. Additional enzymes sequences are found, e.g., in Genbank. Artificially evolved enzymes are also optionally added into a cell in the same manner. In this manner, the cellular machinery and resources of a cell are manipulated to produce unnatural amino acids.

[166] A variety of methods are available for producing novel enzymes for use in biosynthetic pathways or for evolution of existing pathways. For example, recursive recombination, e.g., as developed by Maxygen, Inc. (on the world wide web at

WO 02/085923

PCT/US02/12465

maxygen.com), is optionally used to develop novel enzymes and pathways. See, e.g., Stemmer 1994, "Rapid evolution of a protein in vitro by DNA shuffling," *Nature* Vol. 370 No. 4: Pg. 389-391; and Stemmer, 1994, "DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution," *Proc. Natl. Acad. Sci. USA*. Vol. 91: Pg. 10747-10751. Similarly DesignPath™, developed by Genencor (on the world wide web at genencor.com) is optionally used for metabolic pathway engineering, e.g., to engineer a pathway to create O-methyl-L-tyrosine in *E. coli*. This technology reconstructs existing pathways in host organisms using a combination of new genes, e.g., identified through functional genomics, and molecular evolution and design. Diversa Corporation (on the world wide web at diversa.com) also provides technology for rapidly screening libraries of genes and gene pathways, e.g., to create new pathways.

[167] Typically, the biosynthesis methods of the present invention, e.g., the pathway to create *p*-aminophenylalanine (pAF) from chorismate, do not affect the concentration of other amino acids produced in the cell. For example a pathway used to produce pAF from chorismate produces pAF in the cell while the concentrations of other aromatic amino acids typically produced from chorismate are not substantially affected. Typically the unnatural amino acid produced with an engineered biosynthetic pathway of the present invention is produced in a concentration sufficient for efficient protein biosynthesis, e.g., a natural cellular amount, but not to such a degree as to affect the concentration of the other amino acids or exhaust cellular resources. Typical concentrations produced *in vivo* in this manner are about 10 mM to about 0.05 mM. Once a bacterium is transformed with a plasmid comprising the genes used to produce enzymes desired for a specific pathway and a twenty-first amino acid, e.g., pAF, dopa, O-methyl-L-tyrosine, or the like, is generated, *in vivo* selections are optionally used to further optimize the production of the unnatural amino acid for both ribosomal protein synthesis and cell growth.

Compositions that include proteins with unnatural amino acids

[168] The invention provides compositions of matter, including proteins with at least one unnatural amino acid. The invention also provides compositions of matter that include proteins with at least one unnatural amino acid produced using the compositions and methods of the invention. In one embodiment, the proteins are processed and modified in a cell dependent manner, e.g., phosphorylated, glycosylated, folded, membrane bound, etc.

WO 02/085923

PCT/US02/12465

- [169] In one aspect, the composition optionally includes at least about 10 micrograms, e.g., at least about 50 micrograms, at least about 100 micrograms, at least about 500 micrograms, at least about 1 milligram, or even at least about 10 milligrams or more of the protein, e.g., an amount that can be achieved with in vivo protein production methods (details on recombinant protein production and purification are provided herein). For example, the protein is optionally present in the composition at a concentration of at least about 10 micrograms per liter, at least about 50 micrograms per liter, at least about 100 micrograms per liter, at least about 500 micrograms per liter, at least about 1 milligram per liter, or at least about 10 milligrams per liter of the protein, or more micrograms or protein per liter, e.g., in a cell lysate, pharmaceutical buffer, or other liquid suspension (e.g., in a volume of, e.g., anywhere from about 1nL to about 100L). The production of large quantities (e.g., greater than that typically possible with other methods, e.g., in vitro translation) of a protein including at least one unnatural amino acid is a feature of the invention and is an advantage over the prior art.
- 15 [170] The production of large quantities (e.g., greater than that typically possible with other methods, e.g., in vitro translation) of a protein including at least one unnatural amino acid is a feature of the invention and is an advantage over the prior art. For example, the ability to synthesize large quantities of proteins containing, e.g., heavy atoms, facilitates protein structure determination via, e.g., X-ray crystallography.
- 20 [171] The incorporation of an unnatural amino acid can be done to, e.g., tailor changes in protein structure and/or function, e.g., to change size, acidity, nucleophilicity, hydrogen bonding, hydrophobicity, accessibility of protease target sites, etc. Proteins that include an unnatural amino acid can have enhanced or even entirely new catalytic or physical properties. For example, the following properties are optionally modified by inclusion of
- 25 an unnatural amino acid into a protein: toxicity, biodistribution, structural properties, spectroscopic properties, chemical and/or photochemical properties, catalytic ability, half-life (e.g., serum half-life), ability to react with other molecules, e.g., covalently or noncovalently, and the like. The compositions including proteins that include at least one unnatural amino acid are useful for, e.g., novel therapeutics, diagnostics, catalytic
- 30 enzymes, binding proteins (e.g., antibodies), and e.g., the study of protein structure and function.
- [172] In one aspect of the invention, a composition includes at least one protein with at least one, e.g., at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, or more unnatural amino acids. For a given

WO 02/085923

PCT/US02/12465

protein with more than one unnatural amino acids, the unnatural amino acids can be identical or different (e.g., the protein can include two or more different types of unnatural amino acids, or can include two or more different sites having unnatural amino acids, or both).

- 5 [173] Essentially any protein that includes an unnatural amino acid (and any corresponding coding nucleic acid, e.g., which includes one or more selector codons) can be produced using the compositions and methods herein. No attempt is made to identify the hundreds of thousands of known proteins, any of which can be modified to include one or more unnatural amino acid, e.g., by tailoring any available mutation methods to include
- 10 one or more appropriate selector codon in a relevant translation system. Common sequence repositories for known proteins include GenBank EMBL, DDBJ and the NCBI. Other repositories can easily be identified by searching the internet.
- [174] One preferred class of proteins that can be made using the compositions and methods for in vivo incorporation of unnatural amino acids described herein includes
- 15 therapeutic proteins. Examples of therapeutic and other proteins that can be modified to comprise one or more unnatural include, e.g., Alpha-1 antitrypsin, Angiostatin, Antihemolytic factor, antibodies (further details on antibodies are found below), Apolipoprotein, Apoprotein, Atrial natriuretic factor, Atrial natriuretic polypeptide, Atrial peptides, C-X-C chemokines (e.g., T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10,
- 20 GCP-2, NAP-4, SDF-1, PF4, MIG), Calcitonin, CC chemokines (e.g., Monocyte chemoattractant protein-1, Monocyte chemoattractant protein-2, Monocyte chemoattractant protein-3, Monocyte inflammatory protein-1 alpha, Monocyte inflammatory protein-1 beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262), CD40 ligand, C-kit Ligand, Collagen, Colony stimulating factor (CSF),
- 25 Complement factor 5a, Complement inhibitor, Complement receptor 1, cytokines, (e.g., epithelial Neutrophil Activating Peptide-78, GRO α /MGSA, GRO β , GRO γ , MIP-1 α , MIP-1 δ , MCP-1), Epidermal Growth Factor (EGF), Erythropoietin ("EPO", representing a preferred target for modification by the incorporation of one or more unnatural amino acid), Exfoliating toxins A and B, Factor IX, Factor VII, Factor VIII, Factor X, Fibroblast
- 30 Growth Factor (FGF), Fibrinogen, Fibronectin, G-CSF, GM-CSF, Glucocerebrosidase, Gonadotropin, growth factors, Hedgehog proteins (e.g., Sonic, Indian, Desert), Hemoglobin, Hepatocyte Growth Factor (HGF), Hirudin, Human serum albumin, Insulin, Insulin-like Growth Factor (IGF), interferons (e.g., IFN- α , IFN- β , IFN- γ), interleukins

WO 02/085923

PCT/US02/12465

(e.g., IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, etc.), Keratinocyte Growth Factor (KGF), Lactoferrin, leukemia inhibitory factor, Luciferase, Neurturin, Neutrophil inhibitory factor (NIF), oncostatin M, Osteogenic protein, Parathyroid hormone, PD-ECSF, PDGF, peptide hormones (e.g., Human Growth Hormone), Pleiotropin, Protein A, Protein G, Pyrogenic exotoxins A, B, and C, Relaxin, Renin, SCF, Soluble complement receptor I, Soluble I-CAM 1, Soluble interleukin receptors (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), Soluble TNF receptor, Somatomedin, Somatostatin, Somatotropin, Streptokinase, Superantigens, i.e., Staphylococcal enterotoxins (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), Superoxide

5 dismutase, Toxic shock syndrome toxin (TSST-1), Thymosin alpha 1, Tissue plasminogen activator, Tumor necrosis factor beta (TNF beta), Tumor necrosis factor receptor (TNFR), Tumor necrosis factor-alpha (TNF alpha), Vascular Endothelial Growth Factor (VEGEGF), Urokinase and many others.

[175] Many of these proteins are commercially available (*See*, e.g., the Sigma

15 BioSciences 2002 catalogue and price list), and the corresponding protein sequences and genes and, typically, many variants thereof, are well-known (*see*, e.g., Genebank). Any of them can be modified by the insertion of one or more unnatural amino acid according to the present invention, e.g., to alter the protein with respect to one or more therapeutic property of interest. Examples of therapeutically relevant properties include serum half-

20 life, shelf half-life, stability, immunogenicity, therapeutic activity, detectability (e.g., by the inclusion of reporter groups (e.g., labels or label binding sites) in the unnatural amino acids), reduction of LD-50 or other side effects, ability to enter the body through the gastric tract (e.g., oral availability), or the like.

[176] One class of proteins that can be made using the compositions and methods for in

25 vivo incorporation of unnatural amino acids described herein includes transcriptional and expression activators. Example transcriptional and expression activators include genes and proteins that modulate cell growth, differentiation, regulation, or the like. Expression and transcriptional activators are found in prokaryotes, viruses, and eukaryotes, including fungi, plants, and animals, including mammals, providing a wide range of therapeutic

30 targets. It will be appreciated that expression and transcriptional activators regulate transcription by many mechanisms, e.g., by binding to receptors, stimulating a signal transduction cascade, regulating expression of transcription factors, binding to promoters and enhancers, binding to proteins that bind to promoters and enhancers, unwinding DNA, splicing pre-mRNA, polyadenylating RNA, and degrading RNA.

WO 02/085923

PCT/US02/12465

- [177] One preferred class of proteins of the invention (e.g., proteins with one or more unnatural amino acids) include expression activators such as cytokines, inflammatory molecules, growth factors, their receptors, and oncogene products, e.g., interleukins (e.g., IL-1, IL-2, IL-8, etc.), interferons, FGF, IGF-I, IGF-II, FGF, PDGF, TNF, TGF- α , TGF- β ,
 5 EGF, KGF, SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, and hyalurin/CD44; signal transduction molecules and corresponding oncogene products, e.g., Mos, Ras, Raf, and Met; and transcriptional activators and suppressors, e.g., p53, Tat, Fos, Myc, Jun, Myb, Rel, and steroid hormone receptors such as those for estrogen, progesterone, testosterone, aldosterone, the LDL receptor ligand and corticosterone.
- 10 [178] A variety of other proteins can also be modified to include one or more unnatural amino acid of the invention. For example, the invention can include substituting one or more natural amino acids in one or more vaccine proteins with an unnatural amino acid, e.g., in proteins from infectious fungi, e.g., *Aspergillus*, *Candida* species; bacteria,
 15 particularly *E. coli*, which serves a model for pathogenic bacteria, as well as medically important bacteria such as *Staphylococci* (e.g., *aureus*), or *Streptococci* (e.g., *pneumoniae*); protozoa such as sporozoa (e.g., *Plasmodia*), *rhizopods* (e.g., *Entamoeba*) and flagellates (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); viruses such as (+) RNA viruses (examples include Poxviruses e.g., *vaccinia*; Picornaviruses, e.g. *polio*; Togaviruses, e.g., *rubella*; Flaviviruses, e.g., HCV; and Coronaviruses), (-) RNA viruses
 20 (e.g., Rhabdoviruses, e.g., VSV; Paramyxoviruses, e.g., RSV; Orthomyxoviruses, e.g., influenza; Bunyaviruses; and Arenaviruses), dsDNA viruses (Reoviruses, for example), RNA to DNA viruses, i.e., Retroviruses, e.g., HIV and HTLV, and certain DNA to RNA viruses such as Hepatitis B.
- 25 [179] A variety of enzymes (e.g., industrial enzymes) can also be modified to include one or more unnatural amino acid according to the methods herein, such as amidases, amino acid racemases, acylases, dehalogenases, dioxygenases, diarylpropane peroxidases, epimerases, epoxide hydrolases, esterases, isomerases, kinases, glucose isomerases, glycosidases, glycosyl transferases, haloperoxidases, monooxygenases (e.g., p450s), lipases, lignin peroxidases, nitrile hydratases, nitrilases, proteases, phosphatases,
 30 subtilisins, transaminase, and nucleases.
- [180] Agriculturally related proteins such as insect resistance proteins (e.g., the Cry proteins), starch and lipid production enzymes, plant and insect toxins, toxin-resistance proteins, Mycotoxin detoxification proteins, plant growth enzymes (e.g., Ribulose 1,5-

WO 02/085923

PCT/US02/12465

Bisphosphate Carboxylase/Oxygenase, "RUBISCO"), lipoxygenase (LOX), and Phosphoenolpyruvate (PEP) carboxylase are also suitable targets for unnatural amino acid modification.

5 [181] Genes coding for proteins including at least one unnatural amino acid can be mutagenized using methods well-known to one of skill in the art and described herein under "General Molecular Biology Techniques." For example, a nucleic acid for a protein of interest is mutagenized to include one or more selector codon, providing for insertion of the one or more unnatural amino acids. The present invention includes any such variant, e.g., mutant, versions of any protein, e.g., including at least one unnatural amino acid.

10 [182] Similarly, the present invention also includes corresponding nucleic acids, i.e., any nucleic acid with one or more selector codon that encodes one or more unnatural amino acid.

[183] In one example embodiment, the invention provides compositions that include a Asp112TAG mutant of chloramphenicol acetyltransferase (CAT) produced by the compositions and methods of the invention, where the CAT protein includes at least one 15 unnatural amino acid, e.g., an O-methyl-L-tyrosine, an L-3-(2-naphthyl)alanine, an amino-, isopropyl-, or allyl-containing tyrosine analogue, etc., and the protein is present in the composition at a concentration of at least about 100 micrograms per liter. In another embodiment, the invention provides compositions that include a Tyr163TAG mutant of 20 mouse dihydrofolate reductase (DHFR) where the DHFR protein includes at least one unnatural amino, e.g., an O-methyl-L-tyrosine, an L-3-(2-naphthyl)alanine, an amino-, isopropyl-, or allyl-containing tyrosine analogue, etc., and the protein is present in the composition at a concentration of at least about 100 micrograms per liter.

Making antibodies to proteins comprising unnatural amino acids

25 [184] In one aspect, the present invention provides antibodies to unnatural amino acids and to proteins comprising unnatural amino acids. Antibodies to unnatural amino acids and proteins comprising such unnatural amino acids are useful as purification reagents, e.g., for purifying the proteins and unnatural amino acids of the invention. In addition, the antibodies can be used as indicator reagents to indicate the presence of an unnatural amino 30 acid or protein comprising an unnatural amino acid, e.g., to track the presence or location (e.g., in vivo or in situ) of the unnatural amino acid or protein comprising an unnatural amino acid. It is also, of course, the case that the unnatural amino acid can itself comprise

WO 02/085923

PCT/US02/12465

one or more unnatural amino acids, thereby providing an antibody with one or more property conferred by the one or more unnatural amino acids.

- [185] An antibody of the invention can be a protein comprising one or more polypeptides substantially or partially encoded by immunoglobulin genes or fragments of immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively. A typical immunoglobulin (e.g., antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (VL) and variable heavy chain (VH) refer to these light and heavy chains, respectively.
- [186] Antibodies exist as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce F(ab')₂, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The F(ab')₂ may be reduced under mild conditions to break the disulfide linkage in the hinge region thereby converting the F(ab')₂ dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (*see, Fundamental Immunology*, W.E. Paul, ed., Raven Press, N.Y. (1999), for a more detailed description of other antibody fragments). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such Fab' fragments, etc. may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also optionally includes antibody fragments either produced by the modification of whole antibodies or synthesized *de novo* using recombinant DNA methodologies. Antibodies include single chain antibodies, including single chain Fv (sFv or scFv) antibodies in which a variable heavy and a variable light chain are joined together (directly or through a peptide linker) to form a continuous polypeptide. Antibodies of the invention can be, *e.g.*, polyclonal, monoclonal, chimeric, humanized, single chain, Fab fragments, fragments produced by an Fab expression library, or the like.

WO 02/085923

PCT/US02/12465

[187] In general, antibodies of the invention are valuable, both as general reagents and as therapeutic reagents in a variety of molecular biological or pharmaceutical processes. Methods of producing polyclonal and monoclonal antibodies are available, and can be applied to making the antibodies of the present invention. A number of basic texts

- 5 describe standard antibody production processes, including, e.g., Borrebaeck (ed) (1995) Antibody Engineering, 2nd Edition Freeman and Company, NY (Borrebaeck); McCafferty et al. (1996) Antibody Engineering. A Practical Approach IRL at Oxford Press, Oxford, England (McCafferty), and Paul (1995) Antibody Engineering Protocols Humana Press, Towata, NJ (Paul); Paul (ed.), (1999) Fundamental Immunology, Fifth edition Raven
- 10 Press, N.Y.; Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY; Stites et al. (eds.) Basic and Clinical Immunology (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York, NY; and Kohler and Milstein (1975)
- 15 Nature 256: 495-497.

[188] A variety of recombinant techniques for antibody preparation which do not rely on, e.g., injection of an antigen into an animal have been developed and can be used in the context of the present invention. For example, it is possible to generate and select libraries of recombinant antibodies in phage or similar vectors. See, e.g., Winter et al. (1994)

- 20 "Making Antibodies by Phage Display Technology" Annu. Rev. Immunol. 12:433-55 and the references cited therein for a review. See also, Griffiths and Duncan (1998) "Strategies for selection of antibodies by phage display" Curr Opin Biotechnol 9: 102-8; Hoogenboom et al. (1998) "Antibody phage display technology and its applications" Immunotechnology 4: 1-20; Gram et al. (1992) "in vitro selection and affinity maturation
- 25 of antibodies from a naïve combinatorial immunoglobulin library" PNAS 89:3576-3580; Huse et al. (1989) Science 246: 1275-1281; and Ward, et al. (1989) Nature 341: 544-546.

[189] In one embodiment, antibody libraries can include repertoires of V genes (e.g., harvested from populations of lymphocytes or assembled in vitro) which are cloned for display of associated heavy and light chain variable domains on the surface of filamentous

- 30 bacteriophage. Phage are selected by binding to an antigen. Soluble antibodies are expressed from phage infected bacteria and the antibody can be improved, e.g., via mutagenesis. See e.g., Balint and Larrick (1993) "Antibody Engineering by Parsimonious Mutagenesis" Gene 137:109-118; Stemmer et al. (1993) "Selection of an Active Single Chain Fv Antibody From a Protein Linker Library Prepared by Enzymatic Inverse PCR"

WO 02/085923

PCT/US02/12465

Biotechniques 14(2):256-65; Crameri et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" Nature Medicine 2:100-103; and Crameri and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" BioTechniques 18:194-195.

- 5 [190] Kits for cloning and expression of recombinant antibody phage systems are also known and available, e.g., the "recombinant phage antibody system, mouse ScFv module," from Amersham-Pharmacia Biotechnology (Uppsala, Sweden). Bacteriophage antibody libraries have also been produced for making high affinity human antibodies by chain shuffling (See, e.g., Marks et al. (1992) "By- Passing Immunization: Building High
- 10 Affinity Human Antibodies by Chain Shuffling" Biotechniques 10:779-782. Indeed, antibodies can typically be custom ordered from any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, inc. (www.htibio.com), BMA Biomedicals Ltd (U.K.), Bio.Synthesis, Inc., Research Genetics (Huntsville, Alabama) and many others.
- 15 [191] In certain embodiments, it is useful to "humanize" antibodies of the invention, e.g., where the antibodies are to be administered therapeutically. The use of humanized antibodies tends to reduce the incidence of unwanted immune responses against the therapeutic antibodies (e.g., when the patient is a human). The antibody references above describe humanization strategies. In addition to humanized antibodies, human antibodies
- 20 are also a feature of the invention. Human antibodies consist of characteristically human immunoglobulin sequences. Human antibodies can be produced in using a wide variety of methods (see, e.g., Larrick *et al.*, U.S. Pat. No. 5,001,065, for a review). A general approach for producing human antibodies by trioma technology is described by Ostberg *et al.* (1983), *Hybridoma* 2: 361-367, Ostberg, U.S. Pat. No. 4,634,664, and Engelman *et al.*,
- 25 U.S. Pat. No. 4,634,666.
- [192] A variety of methods of using antibodies in the purification and detection of proteins are known and can be applied to detecting and purifying proteins comprising unnatural amino acids as noted herein. In general, antibodies are useful reagents for ELISA, western blotting, immunochemistry, affinity chromatography methods, SPR, and
- 30 many other methods. The references noted above provide details on how to perform ELISA assays, western blots, surface plasmon resonance (SPR) and the like.
- [193] In one aspect of the invention, antibodies of the invention themselves include unnatural amino acids, providing the antibodies with properties of interest (e.g., improved half-life, stability, toxicity, or the like. Antibodies account for nearly 50% of all

WO 02/085923

PCT/US02/12465

compounds currently in clinical trials (Witttrup, (1999) "Phage on display" Tibtech 17: 423-424 and antibodies are used ubiquitously as diagnostic reagents. Accordingly, the ability to modify antibodies with unnatural amino acids provides an important tool for modifying these valuable reagents.

- 5 [194] For example, there are many applications of MAbs to the field of diagnostics. Assays range from simple spot tests to more involved methods such as the radio-labeled NR-LU-10 MAb from DuPont Merck Co. used for tumor imaging (Rusch et al. (1993) "NR-LU-10 monoclonal antibody scanning. A helpful new adjunct to computed tomography in evaluating non-small-cell lung cancer." J Thorac Cardiovasc Surg 106: 10 200-4). As noted, MAbs are central reagents for ELISA, western blotting, immunochemistry, affinity chromatography methods and the like. Any such diagnostic antibody can be modified to include one or more unnatural amino acid, altering, e.g., the specificity or avidity of the Ab for a target, or altering one or more detectable property, e.g., by including a detectable label (e.g., spectrographic, fluorescent, luminescent, etc.) in 15 the unnatural amino acid.
- [195] One class of valuable antibody reagents are therapeutic Abs. For example, antibodies can be tumor-specific MAbs that arrest tumor growth by Targeting tumor cells for destruction by antibody-dependent cell-mediated cytotoxicity (ADCC) or complement-mediated lysis (CML) (these general types of Abs are sometimes referred to as "magic 20 bullets"). One example is Rituxan, an anti-CD20 MAb for the treatment of Non-Hodgkins lymphoma (Scott (1998) "Rituximab: a new therapeutic monoclonal antibody for non-Hodgkin's lymphoma" Cancer Pract 6: 195-7). A second example relates to antibodies which interfere with a critical component of tumor growth. Herceptin is an anti-HER-2 monoclonal antibody for treatment of metastatic breast cancer, and provides an example of 25 an antibody with this mechanism of action (Baselga et al. (1998) "Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and doxorubicin against HER2/neu overexpressing human breast cancer xenografts [published erratum appears in Cancer Res (1999) 59(8):2020], Cancer Res 58: 2825-31). A third example relates to antibodies for delivery of cytotoxic compounds (toxins, 30 radionuclides, etc.) directly to a tumor or other site of interest. For example, One application Mab is CYT-356, a 90Y-linked antibody that targets radiation directly to prostate tumor cells (Deb et al. (1996) "Treatment of hormone-refractory prostate cancer with 90Y-CYT-356 monoclonal antibody" Clin Cancer Res 2: 1289-97. A fourth application is antibody-directed enzyme prodrug therapy, where an enzyme co-localized to

WO 02/085923

PCT/US02/12465

a tumor activates a systemically-administered pro-drug in the tumor vicinity. For example, an anti-Ep-CAM1 antibody linked to carboxypeptidase A is being developed for treatment of colorectal cancer (Wolfe et al. (1999) "Antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1: in vitro and in vivo studies with prodrugs of methotrexate and the thymidylate synthase inhibitors GW1031 and GW1843" Bioconjug Chem 10: 38-48). Other Abs (e.g., antagonists) are designed to specifically inhibit normal cellular functions for therapeutic benefit. An example is Orthoclone OKT3, an anti-CD3 MAb offered by Johnson and Johnson for reducing acute organ transplant rejection (Strate et al. (1990) "Orthoclone OKT3 as first-line therapy in acute renal allograft rejection" Transplant Proc 22: 219-20. Another class of antibody products are agonists. These Mabs are designed to specifically enhance normal cellular functions for therapeutic benefit. For example, Mab-based agonists of acetylcholine receptors for neurotherapy are under development (Xie et al. (1997) "Direct demonstration of MuSK involvement in acetylcholine receptor clustering through identification of agonist ScFv" Nat. Biotechnol. 15: 768-71. Any of these antibodies can be modified to include one or more unnatural amino acid to enhance one or more therapeutic property (specificity, avidity, serum-half-life, etc.).

[196] Another class of antibody products provide novel functions. The main antibodies in this group are catalytic antibodies such as Ig sequences that have been engineered to mimic the catalytic abilities of enzymes (Wentworth and Janda (1998) "Catalytic antibodies" Curr Opin Chem Biol 2: 138-44. For example, an interesting application involves using the catalytic antibody mAb-15A10 to hydrolyze cocaine *in vivo* for addiction therapy (Mets et al. (1998) "A catalytic antibody against cocaine prevents cocaine's reinforcing and toxic effects in rats" Proc Natl Acad Sci U S A 95: 10176-81).

25 Catalytic antibodies can also be modified to include one or more unnatural amino acid to improve one or more property of interest.

Purifying recombinant proteins comprising unnatural amino acids

[197] Proteins of the invention, e.g., proteins comprising unnatural amino acids, antibodies to proteins comprising unnatural amino acids, etc., can be purified, either partially or substantially to homogeneity, according to standard procedures known to and used by those of skill in the art. Accordingly, polypeptides of the invention can be recovered and purified by any of a number of methods well known in the art, including,

WO 02/085923

PCT/US02/12465

- e.g., ammonium sulfate or ethanol precipitation, acid or base extraction, column chromatography, affinity column chromatography, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, hydroxylapatite chromatography, lectin chromatography, gel electrophoresis and the like. Protein refolding steps can be used, as desired, in making correctly folded mature proteins. High performance liquid chromatography (HPLC), affinity chromatography or other suitable methods can be employed in final purification steps where high purity is desired. In one embodiment, antibodies made against unnatural amino acids (or proteins comprising unnatural amino acids) are used as purification reagents, e.g., for affinity-based purification of proteins comprising one or more unnatural amino acid(s). Once purified, partially or to homogeneity, as desired, the polypeptides are optionally used e.g., as assay components, therapeutic reagents or as immunogens for antibody production.
- [198] In addition to other references noted herein, a variety of purification/protein folding methods are well known in the art, including, e.g., those set forth in R. Scopes, Protein Purification, Springer-Verlag, N.Y. (1982); Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press, Inc. N.Y. (1990); Sandana (1997) Bioseparation of Proteins, Academic Press, Inc.; Bollag et al. (1996) Protein Methods, 2nd Edition Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook Humana Press, NJ; Harris and Angal (1990) Protein Purification Applications: A Practical Approach IRL Press at Oxford, Oxford, England; Harris and Angal Protein Purification Methods: A Practical Approach IRL Press at Oxford, Oxford, England; Scopes (1993) Protein Purification: Principles and Practice 3rd Edition Springer Verlag, NY; Janson and Ryden (1998) Protein Purification: Principles, High Resolution Methods and Applications, Second Edition Wiley-VCH, NY; and Walker (1998) Protein Protocols on CD-ROM Humana Press, NJ; and the references cited therein.
- [199] As noted, those of skill in the art will recognize that, after synthesis, expression and/or purification, proteins can possess a conformation different from the desired conformations of the relevant polypeptides. For example, polypeptides produced by prokaryotic systems often are optimized by exposure to chaotropic agents to achieve proper folding. During purification from, e.g., lysates derived from E. coli, the expressed protein is optionally denatured and then renatured. This is accomplished, e.g., by solubilizing the proteins in a chaotropic agent such as guanidine HCl.

WO 02/085923

PCT/US02/12465

[200] In general, it is occasionally desirable to denature and reduce expressed polypeptides and then to cause the polypeptides to re-fold into the preferred conformation. For example, guanidine, urea, DTT, DTE, and/or a chaperonin can be added to a translation product of interest. Methods of reducing, denaturing and renaturing proteins are well known to those of skill in the art (see, the references above, and Debinski, et al. (1993) J. Biol. Chem., 268: 14065-14070; Kreitman and Pastan (1993) Bioconjug. Chem., 4: 581-585; and Buchner, et al., (1992) Anal. Biochem., 205: 263-270). Debinski, et al., for example, describe the denaturation and reduction of inclusion body proteins in guanidine-DTE. The proteins can be refolded in a redox buffer containing, e.g., oxidized glutathione and L-arginine. Refolding reagents can be flowed or otherwise moved into contact with the one or more polypeptide or other expression product, or vice-versa.

Nucleic acid and polypeptide sequence variants

[201] As described above and below, the invention provides for nucleic acid polynucleotide sequences and polypeptide amino acid sequences, e.g., O-tRNAs and O-RSs, and, e.g., compositions and methods comprising said sequences. Examples of said sequences, e.g., O-tRNAs and O-RSs are disclosed herein. However, one of skill in the art will appreciate that the invention is not limited to those sequences disclosed herein, e.g., the Examples. One of skill will appreciate that the present invention also provides many unrelated sequences with the functions described herein, e.g., encoding an O-tRNA or an O-RS.

[202] One of skill will also appreciate that many variants of the disclosed sequences are included in the invention. For example, conservative variations of the disclosed sequences that yield a functionally identical sequence are included in the invention. Variants of the nucleic acid polynucleotide sequences, wherein the variants hybridize to at least one disclosed sequence, are considered to be included in the invention. Unique subsequences of the sequences disclosed herein, as determined by, e.g., standard sequence comparison techniques, are also included in the invention.

Conservative variations

[203] Owing to the degeneracy of the genetic code, "silent substitutions" (i.e., substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of every nucleic acid sequence which encodes an amino acid. Similarly, "conservative amino acid substitutions," in one or a few amino

WO 02/085923

PCT/US02/12465

acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are also readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the present invention.

- 5 [204] "Conservative variations" of a particular nucleic acid sequence refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or, where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. One of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids
- 10 (typically less than 5%, more typically less than 4%, 2% or 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid. Thus, "conservative variations" of a listed polypeptide sequence of the present invention include substitutions of a small percentage, typically less than 5%,
- 15 more typically less than 2% or 1%, of the amino acids of the polypeptide sequence, with a conservatively selected amino acid of the same conservative substitution group. Finally, the addition of sequences which do not alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional sequence, is a conservative variation of the basic nucleic acid.
- 20 [205] Conservative substitution tables providing functionally similar amino acids are well known in the art. The following sets forth example groups which contain natural amino acids that include "conservative substitutions" for one another.

Conservative Substitution Groups

1	Alanine (A)	Serine (S)	Threonine (T)
2	Aspartic acid (D)	Glutamic acid (E)	
3	Asparagine (N)	Glutamine (Q)	
4	Arginine (R)	Lysine (K)	
5	Isoleucine (I)	Leucine (L)	Methionine (M) Valine (V)
6	Phenylalanine (F)	Tyrosine (Y)	Tryptophan (W)

25

Nucleic Acid Hybridization

[206] Comparative hybridization can be used to identify nucleic acids of the invention, including conservative variations of nucleic acids of the invention, and this comparative

WO 02/085923

PCT/US02/12465

hybridization method is a preferred method of distinguishing nucleic acids of the invention. In addition, target nucleic acids which hybridize to the nucleic acids represented by SEQ ID NO:1-34 under high, ultra-high and ultra-ultra high stringency conditions are a feature of the invention. Examples of such nucleic acids include those with one or a few silent or conservative nucleic acid substitutions as compared to a given nucleic acid sequence.

[207] A test nucleic acid is said to specifically hybridize to a probe nucleic acid when it hybridizes at least $\frac{1}{2}$ as well to the probe as to the perfectly matched complementary target, i.e., with a signal to noise ratio at least $\frac{1}{2}$ as high as hybridization of the probe to the target under conditions in which the perfectly matched probe binds to the perfectly matched complementary target with a signal to noise ratio that is at least about 5x-10x as high as that observed for hybridization to any of the unmatched target nucleic acids.

[208] Nucleic acids "hybridize" when they associate, typically in solution. Nucleic acids hybridize due to a variety of well characterized physico-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* part 1 chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, New York), as well as in Ausubel, *supra*. Hames and Higgins (1995) *Gene Probes 1* IRL Press at Oxford University Press, Oxford, England, (Hames and Higgins 1) and Hames and Higgins (1995) *Gene Probes 2* IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 2) provide details on the synthesis, labeling, detection and quantification of DNA and RNA, including oligonucleotides.

[209] An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see*, Sambrook, *supra* for a description of SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove background probe signal. An example low stringency wash is 2x SSC at 40°C for 15 minutes. In general, a signal to noise ratio of 5x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

WO 02/085923

PCT/US02/12465

- [210] "Stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), *supra*, and in Hames and Higgins, 1 and 2. Stringent hybridization and wash conditions can easily be
- 5 determined empirically for any test nucleic acid. For example, in determining highly stringent hybridization and wash conditions, the hybridization and wash conditions are gradually increased (e.g., by increasing temperature, decreasing salt concentration, increasing detergent concentration and/or increasing the concentration of organic solvents
- 10 such as formalin in the hybridization or wash), until a selected set of criteria are met. For example, the hybridization and wash conditions are gradually increased until a probe binds to a perfectly matched complementary target with a signal to noise ratio that is at least 5x as high as that observed for hybridization of the probe to an unmatched target.
- [211] "Very stringent" conditions are selected to be equal to the thermal melting point
- 15 (T_m) for a particular probe. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the test sequence hybridizes to a perfectly matched probe. For the purposes of the present invention, generally, "highly stringent" hybridization and wash conditions are selected to be about 5° C lower than the T_m for the specific sequence at a defined ionic strength and pH.
- 20 [212] "Ultra high-stringency" hybridization and wash conditions are those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a
- 25 signal to noise ratio of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-high stringency conditions.
- [213] Similarly, even higher levels of stringency can be determined by gradually increasing the hybridization and/or wash conditions of the relevant hybridization assay. For example, those in which the stringency of hybridization and wash conditions are
- 30 increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x, 20X, 50X, 100X, or 500X or more as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to

WO 02/085923

PCT/US02/12465

noise ratio of at least $\frac{1}{2}$ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-ultra-high stringency conditions.

- [214] Nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Unique subsequences

- [215] In one aspect, the invention provides a nucleic acid which comprises a unique subsequence in a nucleic acid selected from the sequences of O-tRNAs and O-RSs disclosed herein. The unique subsequence is unique as compared to a nucleic acid corresponding to any known O-tRNA or O-RS nucleic acid sequence. Alignment can be performed using, *e.g.*, BLAST set to default parameters. Any unique subsequence is useful, *e.g.*, as a probe to identify the nucleic acids of the invention.

- [216] Similarly, the invention includes a polypeptide which comprises a unique subsequence in a polypeptide selected from the sequences of O-RSs disclosed herein. Here, the unique subsequence is unique as compared to a polypeptide corresponding to any of known polypeptide sequence.

- [217] The invention also provides for target nucleic acids which hybridizes under stringent conditions to a unique coding oligonucleotide which encodes a unique subsequence in a polypeptide selected from the sequences of O-RSs wherein the unique subsequence is unique as compared to a polypeptide corresponding to any of the control polypeptides (*e.g.*, parental sequences from which synthetases of the invention were derived, *e.g.*, by mutation). Unique sequences are determined as noted above.

Sequence comparison, identity, and homology

- [218] The terms "identical" or percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (or other algorithms available to persons of skill) or by visual inspection.

- [219] The phrase "substantially identical," in the context of two nucleic acids or polypeptides (*e.g.*, DNAs encoding an O-tRNA or O-RS, or the amino acid sequence of an O-RS) refers to two or more sequences or subsequences that have at least about 60%,

WO 02/085923

PCT/US02/12465

preferably 80%, most preferably 90-95% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm or by visual inspection. Such "substantially identical" sequences are typically considered to be "homologous," without reference to actual ancestry.

- 5 Preferably, the "substantial identity" exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues, or over the full length of the two sequences to be compared.

[220] For sequence comparison and homology determination, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[221] Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally*, Ausubel *et al.*, *infra*).

[222] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be

WO 02/085923

PCT/US02/12465

increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are

5 halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a

10 wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

[223] In addition to calculating percent sequence identity, the BLAST algorithm also

15 performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a

20 reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Defining Polypeptides by Immunoreactivity

[224] Because the polypeptides of the invention provide a variety of new polypeptide

25 sequences (e.g., comprising unnatural amino acids in the case of proteins synthesized in the translation systems herein, or, e.g., in the case of the novel synthetases herein, novel sequences of standard amino acids), the polypeptides also provide new structural features which can be recognized, e.g., in immunological assays. The generation of antisera which specifically bind the polypeptides of the invention, as well as the polypeptides which are

30 bound by such antisera, are a feature of the invention.

[225] For example, the invention includes synthetase proteins that specifically bind to or that are specifically immunoreactive with an antibody or antisera generated against an immunogen comprising an amino acid sequence selected from one or more of (SEQ ID

WO 02/085923

PCT/US02/12465

NO:35-66. To eliminate cross-reactivity with other homologues, the antibody or antisera is subtracted with available synthetases, such as the wild-type *Methanococcus jannaschii* (*M. jannaschii*) tyrosyl synthetase (TyrRS). Where the the wild-type *Methanococcus jannaschii* (*M. jannaschii*) tyrosyl synthetase (TyrRS) corresponds to a nucleic acid, a polypeptide encoded by the nucleic acid is generated and used for antibody/antisera subtraction purposes.

[226] In one typical format, the immunoassay uses a polyclonal antiserum which was raised against one or more polypeptide comprising one or more of the sequences corresponding to one or more of SEQ ID NO:35-66) or a substantial subsequence thereof

(i.e., at least about 30% of the full length sequence provided). The set of potential polypeptide immunogens derived from SEQ ID NO:35-66) are collectively referred to below as "the immunogenic polypeptides." The resulting antisera is optionally selected to have low cross-reactivity against the control synthetase homologues and any such cross-reactivity is removed, e.g., by immunoabsorption, with one or more of the control synthetase homologues, prior to use of the polyclonal antiserum in the immunoassay.

[227] In order to produce antisera for use in an immunoassay, one or more of the immunogenic polypeptides is produced and purified as described herein. For example, recombinant protein can be produced in a recombinant cell. An inbred strain of mice (used in this assay because results are more reproducible due to the virtual genetic identity of the mice) is immunized with the immunogenic protein(s) in combination with a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (see, e.g., Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a standard description of antibody generation, immunoassay formats and conditions that can be used to determine specific immunoreactivity.

Additional references and discussion of antibodies is also found herein and can be applied here to defining polypeptides by immunoreactivity). Alternatively, one or more synthetic or recombinant polypeptide derived from the sequences disclosed herein is conjugated to a carrier protein and used as an immunogen.

[228] Polyclonal sera are collected and titrated against the immunogenic polypeptide in an immunoassay, for example, a solid phase immunoassay with one or more of the immunogenic proteins immobilized on a solid support. Polyclonal antisera with a titer of 10^6 or greater are selected, pooled and subtracted with the control synthetase polypeptides to produce subtracted pooled titrated polyclonal antisera.

WO 02/085923

PCT/US02/12465

[229] The subtracted pooled titrated polyclonal antisera are tested for cross reactivity against the control homologues in a comparative immunoassay. In this comparative assay, discriminatory binding conditions are determined for the subtracted titrated polyclonal antisera which result in at least about a 5-10 fold higher signal to noise ratio for binding of the titrated polyclonal antisera to the immunogenic synthetase as compared to binding to the control synthetase homologues. That is, the stringency of the binding reaction is adjusted by the addition of non-specific competitors such as albumin or non-fat dry milk, and/or by adjusting salt conditions, temperature, and/or the like. These binding conditions are used in subsequent assays for determining whether a test polypeptide (a polypeptide being compared to the immunogenic polypeptides and/ or the control polypeptides) is specifically bound by the pooled subtracted polyclonal antisera. In particular, test polypeptides which show at least a 2-5x higher signal to noise ratio than the control synthetase homologues under discriminatory binding conditions, and at least about a 1/2 signal to noise ratio as compared to the immunogenic polypeptide(s), shares substantial structural similarity with the immunogenic polypeptide as compared to known synthetases, and is, therefore a polypeptide of the invention.

[230] In another example, immunoassays in the competitive binding format are used for detection of a test polypeptide. For example, as noted, cross-reacting antibodies are removed from the pooled antisera mixture by immunoabsorbent with the control polypeptides. The immunogenic polypeptide(s) are then immobilized to a solid support which is exposed to the subtracted pooled antisera. Test proteins are added to the assay to compete for binding to the pooled subtracted antisera. The ability of the test protein(s) to compete for binding to the pooled subtracted antisera as compared to the immobilized protein(s) is compared to the ability of the immunogenic polypeptide(s) added to the assay to compete for binding (the immunogenic polypeptides compete effectively with the immobilized immunogenic polypeptides for binding to the pooled antisera). The percent cross-reactivity for the test proteins is calculated, using standard calculations.

[231] In a parallel assay, the ability of the control proteins to compete for binding to the pooled subtracted antisera is optionally determined as compared to the ability of the immunogenic polypeptide(s) to compete for binding to the antisera. Again, the percent cross-reactivity for the control polypeptides is calculated, using standard calculations. Where the percent cross-reactivity is at least 5-10x as high for the test polypeptides as compared to the control polypeptides and or where the binding of the test polypeptides is

WO 02/085923

PCT/US02/12465

approximately in the range of the binding of the immunogenic polypeptides, the test polypeptides are said to specifically bind the pooled subtracted antisera.

[232] In general, the immunoabsorbed and pooled antisera can be used in a competitive binding immunoassay as described herein to compare any test polypeptide to the immunogenic and/ or control polypeptide(s). In order to make this comparison, the immunogenic, test and control polypeptides are each assayed at a wide range of concentrations and the amount of each polypeptide required to inhibit 50% of the binding of the subtracted antisera to, e.g., an immobilized control, test or immunogenic protein is determined using standard techniques. If the amount of the test polypeptide required for binding in the competitive assay is less than twice the amount of the immunogenic polypeptide that is required, then the test polypeptide is said to specifically bind to an antibody generated to the immunogenic protein, provided the amount is at least about 5-10x as high as for the control polypeptide.

[233] As an additional determination of specificity, the pooled antisera is optionally fully immunosorbed with the immunogenic polypeptide(s) (rather than the control polypeptides) until little or no binding of the resulting immunogenic polypeptide subtracted pooled antisera to the immunogenic polypeptide(s) used in the immunosorption is detectable. This fully immunosorbed antisera is then tested for reactivity with the test polypeptide. If little or no reactivity is observed (i.e., no more than 2x the signal to noise ratio observed for binding of the fully immunosorbed antisera to the immunogenic polypeptide), then the test polypeptide is specifically bound by the antisera elicited by the immunogenic protein.

General Molecular Biology Techniques

[234] General texts which describe molecular biological techniques include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel"). These texts describe mutagenesis, the use of vectors, promoters and many other relevant topics related to, e.g., the generation of genes that include selector codons for production of proteins that include unnatural amino acids, orthogonal tRNAs, orthogonal synthetases, and pairs thereof.

WO 02/085923

PCT/US02/12465

- [235] Various types of mutagenesis are used in the present invention, e.g., to insert selector codons that encode unnatural amino acids in a protein. They include but are not limited to site-directed, random point mutagenesis, homologous recombination (DNA shuffling), mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA or the like. Additional suitable methods include point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, double-strand break repair, and the like. Mutagenesis, e.g., involving chimeric constructs, are also included in the present invention. In one embodiment, mutagenesis can be guided by known information of the naturally occurring molecule or altered or mutated naturally occurring molecule, e.g., sequence, sequence comparisons, physical properties, crystal structure or the like.
- [236] The above texts and examples found herein describe these procedures. Additional information is found in the following publications and references cited within: Ling et al., *Approaches to DNA mutagenesis: an overview*, Anal Biochem, 254(2): 157-178 (1997); Dale et al., *Oligonucleotide-directed random mutagenesis using the phosphorothioate method*, Methods Mol. Biol., 57:369-374 (1996); Smith, *In vitro mutagenesis*, Ann. Rev. Genet., 19:423-462(1985); Botstein & Shortle, *Strategies and applications of in vitro mutagenesis*, Science 229:1193-1201(1985); Carter, *Site-directed mutagenesis*, Biochem. J., 237:1-7 (1986); Kunkel, *The efficiency of oligonucleotide directed mutagenesis*, in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D.M.J. eds., Springer Verlag, Berlin)) (1987); Kunkel, *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Proc. Natl. Acad. Sci. USA 82:488-492 (1985); Kunkel et al., *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Methods in Enzymol., 154, 367-382 (1987); Bass et al., *Mutant Trp repressors with new DNA-binding specificities*, Science 242:240-245 (1988); Methods in Enzymol., 100: 468-500 (1983); Methods in Enzymol., 154: 329-350 (1987); Zoller & Smith, *Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment*, Nucleic Acids Res., 10:6487-6500 (1982); Zoller & Smith, *Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors*, Methods in Enzymol., 100:468-500 (1983); Zoller & Smith, *Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template*, Methods in Enzymol., 154:329-350 (1987); Taylor et

WO 02/085923

PCT/US02/12465

- al., *The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA*, Nucl. Acids Res. 13: 8749-8764 (1985); Taylor et al., *The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA*, Nucl. Acids Res. 13: 8765-8787 (1985); Nakamaye & Eckstein, *Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 14: 9679-9698 (1986); Sayers et al., *Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 16:791-802 (1988); Sayers et al., *Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide*, (1988) Nucl. Acids Res. 16: 803-814; Kramer et al., *The gapped duplex DNA approach to oligonucleotide-directed mutation construction*, Nucl. Acids Res. 12: 9441-9456 (1984); Kramer & Fritz *Oligonucleotide-directed construction of mutations via gapped duplex DNA*, Methods in Enzymol. 154:350-367 (1987); Kramer et al., *Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations*, Nucl. Acids Res. 16: 7207 (1988); Fritz et al., *Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro*, Nucl. Acids Res. 16: 6987-6999 (1988); Kramer et al., *Point Mismatch Repair*, Cell 38:879-887 (1984); Carter et al., *Improved oligonucleotide site-directed mutagenesis using M13 vectors*, Nucl. Acids Res. 13: 4431-4443 (1985); Carter, *Improved oligonucleotide-directed mutagenesis using M13 vectors*, Methods in Enzymol. 154: 382-403 (1987); Eghtedarzadeh & Henikoff, *Use of oligonucleotides to generate large deletions*, Nucl. Acids Res. 14: 5115 (1986); Wells et al., *Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin*, Phil. Trans. R. Soc. Lond. A 317: 415-423 (1986); Nambiar et al., *Total synthesis and cloning of a gene coding for the ribonuclease S protein*, Science 223: 1299-1301 (1984); Sakamar and Khorana, *Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)*, Nucl. Acids Res. 14: 6361-6372 (1988); Wells et al., *Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites*, Gene 34:315-323 (1985); Grundström et al., *Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis*, Nucl. Acids Res. 13: 3305-3316 (1985); Mandecki, *Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis*, Proc. Natl. Acad. Sci. USA, 83:7177-7181 (1986); Arnold, *Protein engineering for unusual environments*, Current Opinion in Biotechnology 4:450-455 (1993); Sieber, et al., *Nature*

WO 02/085923

PCT/US02/12465

Biotechnology, 19:456-460 (2001). W. P. C. Stemmer, *Nature* 370, 389-91 (1994); and, I. A. Lorimer, I. Pastan, *Nucleic Acids Res.* 23, 3067-8 (1995). Additional details on many of the above methods can be found in *Methods in Enzymology* Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis

5 methods.

[237] The present invention also relates to host cells and organisms for the in vivo incorporation of an unnatural amino acid via orthogonal tRNA/RS pairs. Host cells are genetically engineered (e.g., transformed, transduced or transfected) with the vectors of this invention, which can be, for example, a cloning vector or an expression vector. The
10 vector can be, for example, in the form of a plasmid, a bacterium, a virus, a naked polynucleotide, or a conjugated polynucleotide. The vectors are introduced into cells and/or microorganisms by standard methods including electroporation (From et al., *Proc. Natl. Acad. Sci. USA* 82, 5824 (1985), infection by viral vectors, high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads
15 or particles, or on the surface (Klein et al., *Nature* 327, 70-73 (1987)).

[238] The engineered host cells can be cultured in conventional nutrient media modified as appropriate for such activities as, for example, screening steps, activating promoters or selecting transformants. These cells can optionally be cultured into transgenic organisms.

[239] Several well-known methods of introducing target nucleic acids into bacterial cells
20 are available, any of which can be used in the present invention. These include: fusion of the recipient cells with bacterial protoplasts containing the DNA, electroporation, projectile bombardment, and infection with viral vectors (discussed further, below), etc. Bacterial cells can be used to amplify the number of plasmids containing DNA constructs of this invention. The bacteria are grown to log phase and the plasmids within the bacteria
25 can be isolated by a variety of methods known in the art (*see*, for instance, Sambrook). In addition, a plethora of kits are commercially available for the purification of plasmids from bacteria, (*see*, e.g., EasyPrep™, FlexiPrep™, both from Pharmacia Biotech; StrataClean™, from Stratagene; and, QIAprep™ from Qiagen). The isolated and purified plasmids are then further manipulated to produce other plasmids, used to transfect cells or
30 incorporated into related vectors to infect organisms. Typical vectors contain transcription and translation terminators, transcription and translation initiation sequences, and promoters useful for regulation of the expression of the particular target nucleic acid. The vectors optionally comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the cassette in

WO 02/085923

PCT/US02/12465

- eukaryotes, or prokaryotes, or both, (e.g., shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable for replication and integration in prokaryotes, eukaryotes, or preferably both. See, Gilman & Smith, *Gene* 8:81 (1979); Roberts, *et al.*, *Nature*, 328:731 (1987); Schneider, B., *et al.*, *Protein Expr. Purif.* 6435:10 (1995); Ausubel, Sambrook, Berger (*all supra*). A catalogue of Bacteria and Bacteriophages useful for cloning is provided, e.g., by the ATCC, e.g., *The ATCC Catalogue of Bacteria and Bacteriophage* (1992) Gherna *et al.* (eds) published by the ATCC. Additional basic procedures for sequencing, cloning and other aspects of molecular biology and underlying theoretical considerations are also found in Watson *et al.* (1992) *Recombinant DNA Second Edition Scientific American Books, NY*.
- [240] Other useful references, e.g. for cell isolation and culture (e.g., for subsequent nucleic acid isolation) include Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, third edition, Wiley- Liss, New York and the references cited therein; Payne *et al.* (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY; Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL.
- [241] In addition, essentially any nucleic acid (and virtually any labeled nucleic acid, whether standard or non-standard) can be custom or standard ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mrcr@oligos.com), The Great American Gene Company (www.genco.com), ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, CA) and many others.

Pharmaceutical Compositions

- [242] The proteins, e.g., polypeptides, peptides, etc., of the invention (e.g., comprising one or more unnatural amino acid) are optionally employed for therapeutic uses, e.g., in combination with a suitable pharmaceutical carrier. Such compositions, e.g., comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier or excipient includes, but is not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and/or combinations thereof. The formulation is made to suit the mode of administration. In general, methods of administering proteins are well known in the art and can be applied to administration of the polypeptides of the invention.

WO 02/085923

PCT/US02/12465

[243] Therapeutic compositions comprising one or more polypeptide of the invention are optionally tested in one or more appropriate in vitro and/or in vivo animal models of disease, to confirm efficacy, tissue metabolism, and to estimate dosages, according to methods well known in the art. In particular, dosages can be initially determined by activity, stability or other suitable measures of unnatural herein to natural amino acid homologues (e.g., comparison of an EPO modified to include one or more unnatural amino acids to a natural amino acid EPO), i.e., in a relevant assay.

[244] Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. The unnatural amino acid polypeptides of the invention are administered in any suitable manner, optionally with one or more pharmaceutically acceptable carriers. Suitable methods of administering such polypeptides in the context of the present invention to a patient are available, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective action or reaction than another route.

[245] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention.

[246] Polypeptide compositions can be administered by a number of routes including, but not limited to: oral, intravenous, intraperitoneal, intramuscular, transdermal, subcutaneous, topical, sublingual, or rectal means. Unnatural amino acid polypeptide compositions can also be administered via liposomes. Such administration routes and appropriate formulations are generally known to those of skill in the art.

[247] The unnatural amino acid polypeptide, alone or in combination with other suitable components, can also be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

[248] Formulations suitable for parenteral administration, such as, for example, by intraarticular (in the joints), intravenous, intramuscular, intradermal, intraperitoneal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening

WO 02/085923

PCT/US02/12465

agents, stabilizers, and preservatives. The formulations of packaged nucleic acid can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials.

[249] Parenteral administration and intravenous administration are preferred methods of administration. In particular, the routes of administration already in use for natural amino acid homologue therapeutics (e.g., those typically used for EPO, GCSF, GMCSF, IFNs, interleukins, antibodies, and/or any other pharmaceutically delivered protein), along with formulations in current use, provide preferred routes of administration and formulation for the unnatural amino acids of the invention.

[250] The dose administered to a patient, in the context of the present invention, is sufficient to effect a beneficial therapeutic response in the patient over time, or, e.g., to inhibit infection by a pathogen, or other appropriate activity, depending on the application. The dose is determined by the efficacy of the particular vector, or formulation, and the activity, stability or serum half-life of the unnatural amino acid polypeptide employed and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose is also determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular vector, formulation, or the like in a particular patient.

[251] In determining the effective amount of the vector or formulation to be administered in the treatment or prophylaxis of disease (e.g., cancers, inherited diseases, diabetes, AIDS, or the like), the physician evaluates circulating plasma levels, formulation toxicities, progression of the disease, and/or where relevant, the production of anti-unnatural amino acid polypeptide antibodies.

[252] The dose administered, e.g., to a 70 kilogram patient are typically in the range equivalent to dosages of currently-used therapeutic proteins, adjusted for the altered activity or serum half-life of the relevant composition. The vectors of this invention can supplement treatment conditions by any known conventional therapy, including antibody administration, vaccine administration, administration of cytotoxic agents, natural amino acid polypeptides, nucleic acids, nucleotide analogues, biologic response modifiers, and the like.

[253] For administration, formulations of the present invention are administered at a rate determined by the LD-50 of the relevant formulation, and/or observation of any side-effects of the unnatural amino acids at various concentrations, e.g., as applied to the mass and overall health of the patient. Administration can be accomplished via single or divided doses.

WO 02/085923

PCT/US02/12465

- [254] If a patient undergoing infusion of a formulation develops fevers, chills, or muscle aches, he/she receives the appropriate dose of aspirin, ibuprofen, acetaminophen or other pain/fever controlling drug. Patients who experience reactions to the infusion such as fever, muscle aches, and chills are premedicated 30 minutes prior to the future infusions with either aspirin, acetaminophen, or, e.g., diphenhydramine. Meperidine is used for more severe chills and muscle aches that do not quickly respond to antipyretics and antihistamines. Cell infusion is slowed or discontinued depending upon the severity of the reaction.

EXAMPLES

10 Example 1: In vivo incorporation of O-methyl-L-tyrosine

- [255] An orthogonal tRNA/synthetase pair in *E. coli* can be generated by importing a pair from a different organism, if cross-species aminoacylation is inefficient (Y. Kwok, J. T. Wong, Can. J. Biochem. 58, 213-8 (1980)), and the anticodon loop is not a key determinant of synthetase recognition. One such candidate pair is the tyrosyl tRNA/synthetase pair of *Methanococcus jannaschii* (*M. jannaschii*), an archaeobacterium whose tRNA identity elements differ from those of *E. coli* tRNA^{Tyr}, and whose tyrosyl synthetase (TyrRS) lacks an anticodon loop binding domain (B. A. Steer, P. Schimmel, J. Biol. Chem. 274, 35601-6 (1999)). In addition, the *M. jannaschii* TyrRS does not have an editing mechanism (H. Jakubowski, E. Goldman, Microbiol. Rev. 56, 412-29 (1992)), and therefore should not proofread an unnatural amino acid ligated to the tRNA.
- [256] It has been shown that an amber suppressor tRNA derived from the *M. jannaschii* tRNA^{Tyr} is not efficiently aminoacylated by the *E. coli* synthetases, but functions efficiently in protein translation in *E. coli* (L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, J. Am. Chem. Soc. 122, 5010-1 (2000)). Moreover, the *M. jannaschii* TyrRS is orthogonal to *E. coli* tRNAs (B. A. Steer, P. Schimmel, J. Biol. Chem. 274, 35601-6 (1999)), but still efficiently aminoacylates its own suppressor tRNA^{Tyr}_{CUA} (L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, J. Am. Chem. Soc. 122, 5010-1 (2000)). Thus, the *M. jannaschii* tRNA^{Tyr}_{CUA}/TyrRS functions as an orthogonal pair, and can efficiently incorporate tyrosine in response to the amber codon, UAG, in *E. coli*.
- [257] To further reduce recognition of this orthogonal tRNA by *E. coli* synthetases, mutagenesis and selection scheme was performed. For additional details, see United States Patent application Serial No. _____, titled "Methods and Compositions for the

WO 02/085923

PCT/US02/12465

production of orthogonal tRNA-tRNA synthetase pairs," (Attorney docket number 54-000130US, by Schultz et al) filed concurrently with the instant application, the disclosure of which is incorporated in its entirety.

- [258] Briefly, eleven nucleotides of the *M. jannaschii* tRNA^{Tyr}_{CUA} that do not interact directly with the *M. jannaschii* TyrRS were randomly mutated to afford a suppressor tRNA library. This tRNA library was passed through a negative selection that removes tRNAs that are aminoacylated by *E. coli* synthetases, followed by a positive selection for tRNAs that are efficiently aminoacylated by *M. jannaschii* TyrRS. The orthogonality of the resulting suppressor tRNAs was tested by an in vivo complementation assay, based on suppression of an amber stop codon at a nonessential position (Ala184) of the TEM-1 beta-lactamase gene carried on plasmid pBLAM. Aminoacylation of a transformed suppressor tRNA by any endogenous *E. coli* synthetase results in cell growth in the presence of ampicillin. *E. coli* transformed with the *M. jannaschii* tRNA^{Tyr}_{CUA} and pBLAM survive at 55.5 micrograms/mL ampicillin. When the best mutant suppressor tRNA (mtRNA^{Tyr}_{CUA}) selected from the library was expressed, cells survived at only 12.4 micrograms/mL ampicillin. The mutant suppressor mtRNA^{Tyr}_{CUA} contained the following nucleotide substitutions: C17A, U17aG, U20C, G37A, and U47G. For comparison, cells with pBLAM only (in the absence of any suppressor tRNA) survive at 9.7 micrograms/mL ampicillin. When the *M. jannaschii* TyrRS is coexpressed with this mtRNA^{Tyr}_{CUA}, cells survive at 436 micrograms/mL ampicillin. Thus, the mtRNA^{Tyr}_{CUA} is a poorer substrate for the endogenous synthetases than the *M. jannaschii* tRNA^{Tyr}_{CUA} and is still aminoacylated efficiently by the *M. jannaschii* TyrRS.
- [259] To alter the amino acid specificity of the orthogonal *M. jannaschii* TyrRS so that it charges the mtRNA^{Tyr}_{CUA} with a desired unnatural amino acid, a library of TyrRS mutants was generated and screened. Based on the crystal structure of the homologous TyrRS from *Bacillus stearothermophilus* (P. Brick, T. N. Bhat, D. M. Blow, J. Mol. Biol. 208, 83-98 (1988)), five residues (Tyr32, Glu107, Asp158, Ile159 and Leu162) in the active site of *M. jannaschii* TyrRS, which are within 6.5 Å of the para position of the aryl ring of bound tyrosine were mutated. Corresponding residues from a mutant *M. jannaschii* TyrRS (mutTyr RS, for aminoacylation with O-methyl-L-tyrosine) are Tyr³² (Tyr³⁴),

WO 02/085923

PCT/US02/12465

Glu¹⁰⁷ (Asn¹²³), Asp¹⁵⁸ (Asp¹⁷⁶), Ile¹⁵⁹ (Phe¹⁷⁷), and Leu¹⁶² (Leu¹⁸⁰) with *B. stearothermophilus* TyrRS residues in parenthesis.

[260] As described in more detail below, these residues were all initially mutated to alanine to generate an Ala5 TyrRS, which is unable to charge the mtRNA_{CUA}^{Tyr} with

5 tyrosine. This mutant Ala5 TyrRS was used as a template to generate a library of TyrRS mutants in which the five residues were randomly mutated by PCR mutagenesis with doped oligonucleotides.

[261] The *M. jannaschii* TyrRS gene was expressed under the control of *E. coli* GlnRS promoter and terminator in plasmid pBK-JYRS, a pBR322 derived plasmid with

10 kanamycin resistance. Residues Tyr32, Glu107, Asp158, Ile159 and Leu162 were substituted with Ala by site-directed mutagenesis to afford plasmid pBK-JYA5.

Oligonucleotides LW157 5'-GGAATTCATATGGACGAATTGAAATG-3', LW164 5'-GTATTT

TACCACTTGGTTCAAAACCTATMNNAGCAGATTTTTCATCTTTTTCATCTTT

15 TTTTAAAC-3', LW159 5'-TAGGTTTGAACCAAGTGGTAAAATAC-3', LW165

5'-CATTCAGTGTATAATCCTTATCAAGCTGGAAMNNACTTCCATAA

ACATATTTTGCCTTTAAC-3', LW161 5'-TCCAGCTTGATAAGGATTATACA

CTGAATG-3', LW167 5'-CATCCCTCCAACGCAACATCAACGCCMNNATA

ATGMNNMNNATTAACCTGCATTATTGGATAGATAAC-3', LW163 5'-GCGT

20 TGATGTTGCAGTTGGAGGGATG-3', and LW105 5'-AACTGCAGTTATAAT

CCTTTCTAATTGGCTC-3' with NNK (N=A+T+G+C, K=G+T, and M=C+A (Operon,

Alameda, CA) at the mutation sites were used for PCR amplification of the Ala5 TyrRS mutant (pBK-JYA5) and ligated back into the NdeI-PstI-digested pBK-JYA5 to afford the TyrRS library. The ligated vectors were transformed into *E. coli* DH10B competent cells

25 to yield a library of 1.6 X 10⁹ colony forming unit (cfu). The TyrRS genes from 40

randomly picked colonies were sequenced to confirm that there was no base bias at the

randomized NNK positions and no other unexpected mutations. The library was amplified

by maxiprep, and supercoiled DNA was used to transform the selection strain pYC-J17.

[262] A positive selection was then applied that is based on suppression of an amber stop

30 codon at a nonessential position (Asp112) in the chloramphenicol acetyltransferase (CAT) gene (*M. Pastnak, T. J. Magliery, P. G. Schultz, Helvetica Chimica Acta*, 83, 2277-86 (2000)). Cells were grown in media containing the unnatural amino acid and selected for

their survival in the presence of various concentration of chloramphenicol. If a mutant

WO 02/085923

PCT/US02/12465

TyrRS charges the orthogonal $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with any amino acid, either natural or unnatural, the cell produces CAT and survives.

[263] The surviving cells were then grown in the presence of chloramphenicol and in the absence of the unnatural amino acid. Those cells that did not survive, i.e., which encode

5 mutant TyrRSs that charge the orthogonal $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with an unnatural amino acid, were isolated from a replica plate supplemented with the unnatural amino acid. The mutant TyrRS genes were isolated from these cells, recombined in vitro by DNA shuffling, and transformed back into *E. coli* for further rounds of selection.

[264] Seven tyrosine analogues with different functional groups at the para position of
10 the aryl ring (acetyl, amino, carboxyl, isopropyl, methyl, O-methyl and nitro) were used individually in the selections that were performed as follows.

[265] The gene encoding $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ under the control of the lpp promoter and *rmC* terminator was inserted into plasmid pACMD112TAG (a pACYC184 plasmid with a TAG stop codon replacing Asp112 in its CAT gene (M. Pastnak, T. J. Magliery, P. G. Schultz,
15 *Helvetica Chimica Acta*, 83, 2277-86 (2000))) to afford plasmid pYC-J17. Supercoiled DNA encoding the TyrRS library was transformed into *E. coli* DH10B competent cells containing pYC-J17 to yield a library of size greater than 3×10^9 cfu, ensuring complete coverage of the original library. Cells were then plated on minimal media plates containing 1% glycerol and 0.3 mM leucine (GMML) with 17 micrograms/mL
20 tetracycline(Tet), 25 micrograms/mL kanamycin (Kan), 50 micrograms/mL of chloramphenicol (Cm), and 1 mM unnatural amino acid. After incubation at 37 °C for 44 hours, colonies on plates supplied with O-methyl-L-tyrosine were pooled, plasmids were isolated and retransformed into *E. coli* DH10B competent cells containing pYC-J17, and the transformed cells were positively selected on 50 micrograms/mL of Cm. Colonies
25 (96) were individually picked from the plate, diluted into 100 mL of liquid GMML media, and streaked onto two sets of Kan/Tet GMML plates with various concentration of Cm. No O-methyl-L-tyrosine was added to plate set 1 and the concentration of Cm was varied from 10 – 25 micrograms/mL; plate set 2 contained 1mM O-methyl-L-tyrosine and 50 micrograms/mL of Cm. Replicates of colonies that did not grow on 15 micrograms/mL of
30 Cm in plate set 1 were picked from plate set 2. Plasmids containing the TyrRS gene were purified and recombined in vitro by DNA shuffling using Stemmer's protocol (W. P. C. Stemmer, *Nature* 370, 389-91 (1994)) with the exception of 10 mM Mn2+ instead of Mg2+ in the fragmentation reaction (I. A. Lorimer, I. Pastan, *Nucleic Acids Res.* 23, 3067-

WO 02/085923

PCT/US02/12465

- 8 (1995)). The library was then religated into predigested pBK-JYA5 vector to afford a second generation TyrRS library with a typical size of 8×10^8 to 3×10^9 cfu. Thirty randomly selected members from the library were sequenced. The mutagenic rate introduced by DNA shuffling was 0.35%. This library was transformed into the selection
- 5 strain for the next round of selection followed by shuffling. The concentration of Cm in the positive selection and in plate set 2 was raised to 80 micrograms/mL for the second round and 120 micrograms/mL for the third round; the concentration of Cm in plate set 1 was unchanged. After three rounds of DNA shuffling, colonies began to grow on 20 – 25 micrograms/mL Cm in plate set 1, indicating that the TyrRS mutants were accepting
- 10 natural amino acids as substrates. Therefore, the best clone selected after two rounds of DNA shuffling was characterized in detail.
- [266] Following the two rounds of selection and DNA shuffling, a clone (mutant TyrRS (LWJ16)) was evolved whose survival in chloramphenicol was dependent on the addition of 1mM O-methyl-L-tyrosine to the growth media. In the absence of O-methyl-L-
- 15 tyrosine, cells harboring the mutant TyrRS (LWJ16) were not viable on minimal media plates containing 1% glycerol, 0.3 mM leucine (GMML), and 15 micrograms/mL of chloramphenicol. Cells were able to grow on GMML plates with 125 micrograms/mL chloramphenicol in the presence of 1mM O-methyl-L-tyrosine. Similar results were obtained in liquid GMML. As a control, cells with the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ and the inactive Ala5
- 20 TyrRS did not survive at the lowest concentration of chloramphenicol used, either in the presence or absence of 1mM O-methyl-L-tyrosine. This indicates that the growth of cells in chloramphenicol relies on the expression of the mutant TyrRS (LWJ16) and is not a simple nutritional effect of O-methyl-L-tyrosine. Addition of 1mM O-methyl-L-tyrosine itself does not significantly affect the growth rate of *E. coli*.
- 25 [267] To further demonstrate that the observed phenotype is due to the site-specific incorporation of O-methyl-L-tyrosine by the orthogonal $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ /mutant TyrRS (LWJ16) pair in response to an amber stop codon, an O-methyl-L-tyrosine mutant of dihydrofolate reductase (DHFR) was generated and characterized. The third codon of the *E. coli* DHFR gene (a permissive site) was mutated to TAG and a C-terminal His₆ tag was
- 30 added in order to separate the mutant protein from endogenous *E. coli* DHFR. As a control, the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ was coexpressed with the wild type *M. jannaschii* TyrRS, resulting in efficient suppression of the nonsense codon in DHFR with tyrosine. See

WO 02/085923

PCT/US02/12465

Figure 2. When the mutant TyrRS (LWJ16) was expressed in the presence of $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ and 1 mM O-methyl-L-tyrosine in liquid GMML growth media, full length DHFR was also produced and could be purified by Ni affinity chromatography with an isolated yield of 2 mg/liter.

- 5 [268] The yield of purified protein is approximately 26 fold lower in liquid GMML media compare to 2YT rich media. For example, when the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ and wild type *M. jannaschii* TyrRS are coexpressed, the yield of DHFR is 67 mg/L in 2YT and 2.6 mg/L in liquid GMML.
- 10 [269] In the absence of either O-methyl-L-tyrosine, $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ or mutant TyrRS (LWJ16), no DHFR ($< 0.1\%$ by densitometry) was observed by analysis with SDS-polyacrylamide gel electrophoresis and silver staining. See Figure 2. Western analysis further demonstrated that no trace amount of DHFR was produced in the absence of either $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$, mutant TyrRS (LWJ16), or O-methyl-L-tyrosine. See Figure 2.
- 15 [270] The identity of the amino acid inserted in response to the TAG codon was confirmed by mass analysis of both the intact protein and tryptic fragments. The average mass of the intact protein was determined by electrospray ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS). The observed value for the monoisotopic mass from the cluster next to the internal calibrant was 18096.002 Da, which is within 5 ppm of the theoretical mass of 18095.908 Da and clearly demonstrates
- 20 the incorporation of O-methyl-L-tyrosine.
- [271] For this experiment a DHFR mutant lacking the C-terminal His tag was used and purified by methotrexate affinity chromatography. In the mutant protein, the third codon was changed to TAG, and the fourth codon was changed from CTG to ATG to improve the amber suppression efficiency, resulting in a Leu4Met mutation.
- 25 [272] This result also indicates that other endogenous *E. coli* synthetases do not utilize O-methyl-L-tyrosine as a substrate. Liquid chromatography tandem mass spectrometry of tryptic digests was carried out to confirm the sequence of the N-terminal peptide. An example of a tandem MS spectrum is shown in Figure 3. The doubly charged precursor ion at 691.5 Da, corresponding to the N-terminal tryptic peptide MIY*MIAALAVDR,
- 30 was selected and fragmented in an ion trap mass spectrometer (ITMS). The fragment ion masses could be unambiguously assigned as shown in Figure 3, confirming the site-specific incorporation of O-methyl-L-tyrosine. Neither the protein mass spectra nor the

WO 02/085923

PCT/US02/12465

tryptic peptide maps gave any indications of the incorporation of tyrosine or other amino acids in place of O-methyl-L-tyrosine – from the signal-to-noise ratio of the protein mass spectra a minimum 95% incorporation purity for O-methyl-L-tyrosine was obtained.

[273] Taken together, the cell growth, protein expression and mass spectrometry

- 5 experiments demonstrate that the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ /mutant TyrRS (LWJ16) orthogonal pair is capable of selectively inserting O-methyl-L-tyrosine into proteins in response to the amber codon with a fidelity rivaling that of the natural amino acids.

[274] Analysis of the sequence of the mutant TyrRS (LWJ16) that charges the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with O-methyl-L-tyrosine revealed 12 nucleotide changes, two of which were

- 10 silent. The ten nonsilent mutations resulted in the following amino acid residue substitutions relative to wild type TyrRS: Tyr32, which hydrogen bonds to the aryl oxygen atom of the native substrate tyrosine, was mutated to Gln; Asp158, which hydrogen bonds to the hydroxyl group of tyrosine, was mutated to Ala; Glu107, which hydrogen bonds to Asp158, was mutated to Thr; and Leu162, which is located at the bottom of the binding
15 pocket, was mutated to Pro. Based on the x-ray crystal structure of the homologous *B. stearothermophilus* TyrRS, it can be speculated that loss of the hydrogen-bonding network between Tyr32, Asp158 and substrate tyrosine should disfavor binding of tyrosine to the mutant TyrRS (LWJ16). Indeed, mutation of Asp176 (which corresponds to Asp158 in *M. jannaschii*) of *B. stearothermophilus* TyrRS yields inactive enzyme (G. D. P. Gray, H.
20 W. Duckworth, A. R. Fersht, *FEBS* 318, 167-71 (1993)). At the same time, the Asp158Ala and Leu162Pro mutations create a hydrophobic pocket that allows the methyl group of O-methyl-L-tyrosine to extend further into the substrate-binding cavity. Other important catalytic residues in the active site, which bind to the ribose or the phosphate group of the adenylate, were unchanged after two rounds of DNA shuffling. Detailed
25 analysis of these mutations awaits the three-dimensional structure of the mutant TyrRS (LWJ16).

- [275] Kinetics of adenylate formation of O-methyl-L-tyrosine and tyrosine with ATP catalyzed by the mutant TyrRS (LWJ16) were analyzed in vitro using a pyrophosphate-exchange assay. The mutant TyrRS (LWJ16) gene with six histidines at its C-terminus
30 was cloned into plasmid pQE-60 (Qiagen, CA) to afford plasmid pQE-mJYRS. Protein was purified by immobilized metal affinity chromatography according to manufacture's protocol (Qiagen, CA). Pyrophosphate (PPi) exchange was carried out at 37 °C in a reaction mixture containing 100 mM TrisHCl (pH7.5), 10 mM KF, 5 mM MgCl₂, 2 mM

WO 02/085923

PCT/US02/12465

ATP, 2 mM NaPPi, 0.1 mg/mL bovine serum albumin, approximately 0.01 microCi/mL [32P]NaPPi, and various concentrations of tyrosine or O-methyl-L-tyrosine. Reactions were initiated with the addition of the purified mutant TyrRS (LWJ16), and aliquots were periodically taken and quenched with 0.2 M NaPPi, 7% perchloric acid, and 2% activated charcoal. The charcoal was filtered and washed with 10 mM NaPPi (pH2), then measured by scintillation counting to determine the 32P levels in charcoal-adsorbed ATP. Values of kcat and Km were calculated by direct fitting of the Michaelis-Menten equation using nonlinear regression analysis.

[276] Table 1. Kinetic parameters for the mutant TyrRS (LWJ16) toward tyrosine and O-methyl-L-tyrosine measured by pyrophosphate exchange assay.

Amino acid	Kcat (10 ⁻³ s ⁻¹)	Km (μM)	kcat/Km (s ⁻¹ M ⁻¹)
O-methyl-L-tyrosine	14 ± 1	443 ± 93	32
L-tyrosine	1.8 ± 0.2	5833 ± 902	0.31

[277] The results of this analysis are shown in Table 1. The Km for tyrosine (5833 μM) is approximately 13 fold higher than that for O-methyl-L-tyrosine, and the kcat for tyrosine (1.8 X 10⁻³ s⁻¹) is 8 fold down relative to that for O-methyl-L-tyrosine. Thus the value of kcat/Km of the mutant TyrRS (LWJ16) for O-methyl-L-tyrosine is about 100 fold higher than that of tyrosine. The physiological concentration of tyrosine in E. coli is about 80 μM, which is far below Km value (5833 μM) of the mutant TyrRS (LWJ16) for tyrosine. Presumably, the concentration of O-methyl-L-tyrosine in cells is comparable or greater than the Km (443 μM).

20 Example 2: In vivo incorporation of L-3-(2-naphthyl)alanine

[278] The site-specific incorporation of a second unnatural amino acid, L-3-(2-naphthyl)-alanine into proteins in E. coli was accomplished. This result shows that this overall scheme is applicable to a host of amino acids. No synthetase specific for L-3-(2-naphthyl)-alanine were selected from the mutant TyrRS library produced in Example 1, described above.

[279] An amber stop codon and its corresponding orthogonal amber suppressor tRNA, mRNA_{CUA}^{Tyr}, were selected to encode the unnatural amino acid (Wang, L.; Schultz, P. G. *Chem. Biol.* 8, 883-890 (2001)). The *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) was used as the starting point for the generation of an orthogonal synthetase with unnatural

WO 02/085923

PCT/US02/12465

amino acid specificity. This TyrRS does not aminoacylate any endogenous *E. coli* tRNAs (Steer, B. A.; Schimmel, P. J. *Biol. Chem.*, 274, 35601-35606 (1999)), but aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with tyrosine (Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.*, 122, 5010-5011 (2000)). L-3-(2-naphthyl)-alanine was chosen for this study since it represents a significant structural perturbation from tyrosine and may have novel packing properties.

[280] To change the amino acid specificity of the TyrRS so that it charges the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with L-3-(2-naphthyl)-alanine and not any common 20 amino acids, a library of *M. jannaschii* TyrRS mutants was generated and screened. Based on an analysis of the crystal structure of the homologous TyrRS from *Bacillus stearothermophilus* (Brick, P.; Bhat, T. N.; Blow, D. M. *J. Mol. Biol.*, 208, 83-98 (1989)) five residues (Tyr³², Asp¹⁵⁸, Ile¹⁵⁹, Leu¹⁶², and Ala¹⁶⁷) in the active site of *M. jannaschii* TyrRS that are within 7 Å of the *para* position of the aryl ring of tyrosine were mutated. To reduce the wild-type synthetase contamination in the following selection, these residues (except Ala¹⁶⁷) were first all mutated to alanine. The resulting inactive Ala¹⁶⁷ TyrRS gene was used as a template for polymerase chain reaction (PCR) random mutagenesis with oligonucleotides bearing random mutations at the corresponding sites.

[281] The mutant TyrRS library was first passed through a positive selection based on suppression of an amber stop codon at a nonessential position (Asp112) in the chloramphenicol acetyltransferase (CAT) gene. Cells transformed with the mutant TyrRS library and the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ gene were grown in minimal media containing 1 mM L-3-(2-naphthyl)-alanine and 80 µg/mL chloramphenicol. Cells can survive only if a mutant TyrRS aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with either natural amino acids or L-3-(2-naphthyl)-alanine. The surviving cells were then grown in the presence of chloramphenicol and the absence of the unnatural amino acid. Those cells that did not survive must encode a mutant TyrRS that charges the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with L-3-(2-naphthyl)-alanine, and were picked from a replica plate supplied with the unnatural amino acid. After three rounds of positive selection followed by a negative screen, four mutant TyrRS's were characterized using an in vivo assay based on the suppression of the Asp112TAG codon in the CAT gene. In the absence of L-3-(2-naphthyl)-alanine, cells expressing the selected TyrRS and the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ survived in 25 to 35 µg/mL chloramphenicol on minimal media plates containing 1% glycerol and 0.3 mM leucine (GMML plate); in the presence of L-3-(2-

WO 02/085923

PCT/US02/12465

naphthyl)-alanine, cells survived in 100 to 120 µg/mL chloramphenicol on GMMML plates. Compared to the IC₅₀ value in the absence of any TyrRS (4 µg/mL chloramphenicol), these results indicate that the selected TyrRS's accept L-3-(2-naphthyl)-alanine, but also still charge natural amino acids to some degree.

- 5 [282] To further reduce the activity of the mutant TyrRS toward natural amino acids, one round of DNA shuffling was carried out using the above four mutant genes as templates. The resulting mutant TyrRS library was passed through two additional rounds of positive selections and negative screens. One mutant TyrRS (SS12-TyrRS) was evolved, whose activity for natural amino acids was greatly reduced (IC₅₀=9 µg/mL chloramphenicol)
- 10 while its activity toward L-3-(2-naphthyl)-alanine was enhanced (IC₅₀=150 µg/mL chloramphenicol).
- [283] The results of the above described in vivo CAT assays using various mutant TyrRS are shown in Table 2. A pYC-J17 plasmid was used to express the mtRNA^{Tyr}_{CUA} gene and the chloramphenicol acetyltransferase gene with an amber stop codon at Asp112. A
- 15 pBK plasmid was used to express TyrRS, and was cotransformed with pYC-J17 into *E. coli* DH10B. Cell survival on GMMML plates was titrated in the presence of different concentrations of chloramphenicol.

[284] Table 2. In vivo chloramphenicol acetyltransferase assay of mutant TyrRS.

Mutant TyrRS	IC ₅₀ (µg/mL of chloramphenicol)	
	No L-3-(2-naphthyl)-Ala	Add L-3-(2-naphthyl)-Ala
no TyrRS	4	4
wt TyrRS	240	240
After selection		
S1-TyrRS	30	120
S2-TyrRS	30	120
S3-TyrRS	25	110
S4-TyrRS	35	100
After DNA shuffling		
SS12-TyrRS	9	150

- 20 [285] An L-3-(2-naphthyl)-alanine mutant of mouse dihydrofolate reductase (DHFR) was generated and characterized to confirm the ability of the mtRNA^{Tyr}_{CUA}/SS12-TyrRS pair to site-specifically incorporate L-3-(2-naphthyl)-alanine in response to an amber stop codon. The Tyr163 codon of the mouse DHFR gene was mutated to TAG, and a His6 tag
- 25 was added to the COOH-terminus of DHFR to facilitate protein purification using Ni²⁺ affinity chromatography. As a positive control, wild-type *M. jannaschii* TyrRS was

WO 02/085923

PCT/US02/12465

coexpressed with the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ resulting in efficient suppression of the TAG codon with tyrosine (Figure 4). When SS12-TyrRS was coexpressed with the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ in the presence of 1 mM L-3-(2-naphthyl)-alanine, full-length mouse DHFR was also generated (with yield of 2.2 mg/L in liquid GMML minimal medium). In the absence of either L-3-(2-naphthyl)-alanine, $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$, or SS12-TyrRS, no full length DHFR was produced. A penta-His antibody was used to detect the His6 tag at the COOH-terminus of DHFR in a Western blot. No DHFR could be detected in the absence of each of the above three components.

[286] Tryptic digests of the L-3-(2-naphthyl)-alanine mutant of mouse DHFR were analyzed by MALDI FT-ICR and liquid chromatography tandem mass spectrometry to confirm unambiguously the incorporation of L-3-(2-naphthyl)-alanine. The peptide map of the internally calibrated digest shows a major peak at 1867.962, which is within 3.5 ppm of the theoretical mass of the tryptic peptide LLPEX*⁺TGVLSEVQEEK where X* represents the L-3-(2-naphthyl)-alanine residue (Pro164 was mutated to Thr to improve the amber suppression efficiency). Further, the interpreted tandem mass spectrum of precursor ion at m/z 934.5, which corresponds to the doubly charged ion of the peptide of interest is shown in Figure 5. The sequence information gleaned from the spectrum clearly demonstrates the site-specific incorporation of L-3-(2-naphthyl)-alanine into the protein. Neither peptide maps nor LC MS/MS runs produced any indication of mutants in which the L-3-(2-naphthyl)-alanine residue is substituted by other amino acids. The signal-to-noise ratio of more than 1500 observed in the peptide maps shows a fidelity in the incorporation of L-3-(2-naphthyl)-alanine of better than 99.8%.

[287] The evolved SS12-TyrRS has the following mutations: Tyr32→Leu32, Asp158→Pro158, Ile159→Ala159, Leu162→Gln162, and Ala167→Val167. Corresponding residues from *B. stearothermophilus* are Tyr³² (Tyr³⁴), Asp¹⁵⁸ (Asp¹⁷⁶), Ile¹⁵⁹ (Phe¹⁷⁷), Leu¹⁶² (Leu¹⁸⁰), and Ala¹⁶⁷ (Gln¹⁸⁹) with *B. stearothermophilus* TyrRS residues in parenthesis.

[288] Based on the crystal structure of the homologous *B. stearothermophilus* TyrRS, the mutations of Tyr32→Leu32 and Asp158→Pro158 probably result in the loss of hydrogen bonds between Tyr32, Asp158, and the native substrate tyrosine, thus disfavoring the binding of tyrosine to SS12-TyrRS. Most residues are mutated to amino acids with hydrophobic side chains, which are expected to favor binding of L-3-(2-naphthyl)-alanine.

WO 02/085923

PCT/US02/12465

[289] In summary, the cell growth, protein expression, and mass spectrometry experiments demonstrate that the mtRNA_{CUA}^{Tyr}/SS12-TyrRS pair is capable of selectively inserting L-3-(2-naphthyl)-alanine into proteins in response to the amber codon with fidelity rivaling that of the natural amino acids.

5 **Example 3: In vivo incorporation of amino-, isopropyl-, or allyl-containing tyrosine analogues**

[290] A FACS based screening system was used to rapidly evolve three highly selective synthetase variants that accept amino-, isopropyl-, or allyl-containing tyrosine analogues. The system included a multipurpose reporter plasmid used for application of both positive and negative selection pressure and for the facile and quantitative evaluation of synthetase activity. A chloramphenicol acetyl transferase (CAT) marker allowed positive selection for activity of the *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS). A T7 polymerase/GFP reporter system allowed assessment of synthetase activity within cells grown in both the presence and absence of an unnatural amino acid. Fluorescence activated cell sorting (FACS) was used to screen against synthetase variants that accept natural amino acids, while visual and fluorimetric analyses were to assess synthetase activity qualitatively and quantitatively, respectively.

[291] **Design of an amplifiable fluorescence reporter system.** Efforts to develop a versatile screening system for the assessment of synthetase activity in living cells initially arose out of a desire for a greater degree of control over the selective pressure applied to populations of synthetase variants, especially negative selective pressure. As the system was to be used to assess the activities of large numbers of synthetase variants, a reporter was sought that would be amenable to high-throughput screening. In addition, a reporter that would allow for facile qualitative and quantitative evaluation of synthetase activity was desired. To meet these requirements, a fluorescence-based screen was designed. The system was based on the synthetase-dependent production of GFPuv, a variant of the green fluorescent protein that has been optimized for expression in *E. coli* (Cramer, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P., *Nature Biotechnol.* 1996, 14, 315-319). This fluorophore is amenable to use in FACS and fluorimetry, as well as visual inspection on plates and in liquid culture. The system was designed such that synthetase-dependent suppression of selector, e.g., amber nonsense codons would result in the production of a fluorescence signal. In order to maximize the sensitivity of the reporter, it was made

WO 02/085923

PCT/US02/12465

amplifiable by placement of the amber codons within the gene for T7 RNA polymerase, which was designed to drive expression of the GFPuv reporter gene in analogy to other amplifiable intracellular reporter systems (Lorincz, M., Roederer, M., Diwu, Z., Herzenberg, L.A., Nolan, G.P. *Cytometry*, 1996, 24, 321-329; Zlokarnik, G., Negulescu, P.A., Knapp, T.E., Mere, L., Burres, N., Feng, L., Whitney, M., Roemer, K. & Tsien, R.Y., *Science*, 1998, 279, 84-88). The T7 RNA polymerase gene was placed under control of the arabinose promoter in order to allow facile optimization of the production of the RNA transcript for amber codon-containing T7 RNA polymerase.

- [292] **Optimization of the T7 RNA polymerase/GFPuv reporter system.** A medium-copy reporter plasmid, pREP, was designed to express amber-containing T7 RNA polymerase variants under control of the arabinose promoter and the GFPuv gene under control of the T7 promoter (Figure 6a). A series of twelve T7 RNA polymerase variants, designed to optimize synthetase-dependent fluorescence enhancement (Figure 6b), were inserted into pREP to create plasmids pREP(1-12). All variants contained an N-terminal leader sequence of seven amino acids (MTMITVH) and 1-3 amber stop codons (TAG). Variants 1-3 contained one, two, and three amber stop codons, respectively, substituted for the original methionine at position one (M1), just downstream of the leader sequence. Variants 4-9 contained an amber codon substituted for D10, R96, Q107, A159, Q169, or Q232, respectively, which were predicted to be located in loop regions of the structure (Jeruzalmi, D. & Steitz, T.A., *EMBO J.*, 1998, 17, 4101-4113). Variants 10-12 contained amber stop codons substituted at positions M1 and either Q107, A159, or Q232, respectively. Plasmid constructs were evaluated by fluorimetry and flow cytometry of live cells for fluorescence enhancement using a compatible plasmid containing the orthogonal glutamyl-tRNA synthetase and Glutamine tRNA_{CUA} from *S. cerevisiae*. Plasmids pREP(1-12) were found to provide varying levels of synthetase-dependent fluorescence enhancement, with the best construct, pREP(10) exhibiting 220-fold greater fluorescence by fluorimetry (Figure 6c) and ~400-fold greater median fluorescence by cytometry (Figure 6d) in cells containing the wild type synthetase versus an inactive mutant. Substitution of a variety of functional groups at positions corresponding to the amber codons within pREP(10) demonstrate that position 107 within T7 RNA polymerase is highly permissive.

[293] **Construction of a multipurpose reporter plasmid.** In order to construct a multipurpose plasmid to be used both for selecting and screening variants of a *M. jannaschii* TyrRS, plasmid pREP(10) was combined with plasmid pYC-J17 (Wang, L.,

WO 02/085923

PCT/US02/12465

- Brock, A., Herberich, B. & Schultz, P.G., *Science*, 2001, 292, 498-500) to obtain pREP/YC-JYCUA (Figure 7b). Plasmid pREP/YC-JYCUA was assayed for function with a compatible plasmid expressing a variant of *M. jannaschii* TyrRS (pBK-mJYRS; Wang, L., Brock, A., Herberich, B. & Schultz, P.G., *Science*, 2001, 292, 498-500) selective for incorporating *O*-Methyl-Tyrosine (OMY). Cells containing pREP/YC-JYCUA and pBK-mJYRS, grown in the presence of OMY, exhibited a chloramphenicol (Cm) IC₅₀ value of 120 micrograms/ml, identical to that obtained using plasmid pYC-J17, and a fluorescence enhancement of 330-fold for cells grown in the presence versus the absence of OMY, as measured by fluorimetry.
- 10 [294] **Evolution of the substrate specificity of the *M. jannaschii* tyrosyl-tRNA synthetase.** Results have shown that the amino acid side chain binding pocket of the *M. jannaschii* TyrRS can be evolved to selectively accommodate chemical groups other than the phenol side chain of tyrosine (Wang, L., Brock, A., Herberich, B. & Schultz, P.G., *Science*, 2001, 292, 498-500; Wang, L., Brock, A. & Schultz, P.G. *J. Am. Chem. Soc.* 15 2002, 124, 1836-1837). We sought to further explore the generality of unnatural amino acid accommodation by *M. jannaschii* TyrRS by challenging the enzyme to accept four new functionalities: *p*-Isopropyl-Phenylalanine (pIF), *p*-Amino-Phenylalanine (pAF), *p*-Carboxyl-Phenylalanine (pCF), or *O*-Allyl-Tyrosine (OAT) (Figure 7b). A library of *M. jannaschii* TyrRS variants containing randomizations at positions Y32, E107, D158, I159, 20 and L162 (Wang, L., Brock, A., Herberich, B. & Schultz, P.G., *Science*, 2001, 292, 498-500), residues thought to form the binding pocket for the *para* position of the tyrosyl ring, was introduced into cells containing plasmid pREP/YC-JYCUA. These cells, encompassing a library diversity of ~10⁹, were used to begin four evolution experiments to identify synthetase variants selective for pIF, pAF, pCF, or OAT (Figure 7c). Two cycles 25 of positive selection were carried out by allowing the cell cultures to grow to saturation in the presence of Cm and one of the four unnatural amino acids. Cell aliquots were removed following the second cycle of positive selection and used to inoculate a new culture containing no added amino acid or Cm, and the culture was again allowed to grow to saturation. At this point, cells that fluoresce are likely to contain synthetase variants 30 that can accept one of the 20 natural amino acids. Approximately 10⁸ cells from each line were subjected to negative screening using FACS in order to eliminate natural amino acid-accepting synthetase variants. The non-fluorescent cells were collected and amplified through growth to saturation. These amplified cells were used to inoculate a new culture for a final cycle of positive selection in liquid culture containing unnatural amino acid and

WO 02/085923

PCT/US02/12465

Cm. Following growth to saturation, each population of cells was plated on media containing 0, 30, 60, or 100 micrograms/mL Cm and either 0 or 1 mM of the appropriate unnatural amino acid.

5 [295] **Identification and characterization of evolved synthetase variants.** Cm plates supplemented with *pIF*, *pAF*, and *OAT* produced 10–100-fold greater numbers of fluorescent colonies than plates containing no added amino acid. In contrast, plates for the *pCF* population produced the same number of fluorescent colonies with or without addition of *pCF*. The ten largest fluorescent colonies were picked for each of the *pIF*, *pAF*, and *OAT* populations from unnatural amino acid-containing plates and grown to 10 saturation in liquid media with or without added unnatural amino acid. A qualitative assessment of fluorescence production was made visually with the use of a hand-held long-wavelength ultraviolet lamp (Figure 8a).

[296] Synthetase variants corresponding to clones producing significant differences in fluorescence were sequenced. All ten clones from the *pIF* and *pAF* populations had 15 identical sequences, while three different clones were identified from the *OAT* population. Amino acid changes occurred within the five randomized sites in all clones, with the exception of two additional substitutions within the *pIF*-tRNA synthetase (*pIF*-RS) variant. The activities of the different clones were quantitatively assessed. Fluorescence was measured fluorimetrically for cells grown in liquid culture in the presence or absence 20 of unnatural amino acid (Figure 8b). The Cm IC₅₀s were determined by plating the cells on varying concentrations of Cm in the presence or absence of unnatural amino acid (Figure 8c).

[297] A myoglobin gene containing an amber codon in the fourth position was used to assess the production of unnatural amino acid-containing protein. The gene was expressed 25 in cells, using the *pIF*-RS, *pAF*-RS, or *OMY*-RS variant, respectively, in either the presence or absence of *pIF*, *pAF*, or *OAT* (Figure 8d). Protein yields were comparable for all three variants, ranging from 1–2 milligrams of protein per liter of unnatural amino acid-containing cell culture. In contrast, protein production was virtually undetectable in cultures grown in the absence of unnatural amino acid. Proteins were analyzed by 30 electrospray mass spectrometry, giving masses of 18457.40 ± 0.81 (18457.28 expected) for the *pIF*-containing protein, 18430.30 ± 0.27 (18430.21 expected) for the *pAF*-containing protein. Activity measurements obtained using the Cm IC₅₀, fluorimetry, and protein expression analyses correlated well, however the activity of the *pIF*-RS appears to be somewhat underestimated by fluorimetry. As compared to other assays, the

WO 02/085923

PCT/US02/12465

disproportionately low fluorimetry measurement for the *pIF*-RS variant, shows that T7 RNA polymerase may be partially destabilized upon incorporation of the *pIF* analogue, despite the apparent permissivity of the amber positions within the reporter.

5 [298] **Utility of the multipurpose reporter system.** The reporter system described here allows the use of a single multipurpose plasmid for both positive selection and negative screening, obviating the need to shuttle plasmids between alternating rounds of positive and negative selection. A total of only three rounds of positive selection and one round of negative screening were required to enable the identification of synthetase variants that selectively accept desired unnatural amino acids. These features allow evolution
10 experiments to be carried out in a matter of days. The screening system can be used to readily identify active synthetase variants using agar plates containing unnatural amino acid and to individually assay the amino acid specificity of the variants.

[299] As described above, the T7 RNA polymerase/GFP system can be used to quantitatively compare the activities of synthetase variants. The availability of the three
15 *OAT*-RS clones described here and a different *OAT*-RS clone derived independently from the same library using a positive/negative selection based on CAT and barnase (Table 2) allows the possibility of comparing the two different evolution systems in terms of the synthetase variants resulting from each (Figure 9). This analysis reveals that the three clones derived from positive selection and negative screening exhibit slightly lower levels
20 of fluorescence in the presence of *OAT*, but ~10-fold lower background levels in the absence of the unnatural amino acid. The fluorescence enhancement for cells grown in the presence versus the absence of the unnatural amino acid is thus about 6-fold higher for cells expressing *OAT*-RS(1) from selection and screening than for cells expressing the *OAT*-RS clone derived from positive/negative selection using barnase. Although it is not
25 clear whether this example is representative, these data suggest that the T7 RNA polymerase/GFP system may allow more stringency in selecting against synthetase variants that are promiscuous towards natural amino acid substrates. However, the fluorescence enhancement for cells grown in the presence versus the absence of an unnatural amino acid is expected to represent a lower limit for the fidelity of unnatural
30 amino acid incorporation, as competition of unnatural amino acids for being bound by an evolved synthetase variant would reduce binding of natural amino acids. Moreover, although high fidelity is clearly desirable, there is likely to be a trade-off between fidelity and overall synthetase activity, which may depend on the desired application.

WO 02/085923

PCT/US02/12465

- [300] **Generality of aminoacyl tRNA synthetase evolution.** Previous results and those presented here demonstrate that the amino acid side chain binding pocket of the *M. jannaschii* TyrRS is quite malleable. The enzyme can be evolved to accommodate a variety of functionalities in place of the phenol side chain of tyrosine and can do so with high selectivity. In this application it was demonstrated that enzyme can be evolved to accommodate an amine, isopropyl, or allyl ether functionality at the *para* position of the tyrosine ring, instead of hydroxyl.
- [301] **Plasmid Construction.** Plasmid pREP (Figure 6a) was constructed by insertion of a *Bam*HI/*Apa*LI overlap PCR fragment containing the T7 RNA polymerase gene upstream of an *rmb* transcription termination region, followed by an *Apa*LI/*Ahd*II overlap PCR fragment containing the *araC* gene and *ara* promoter region from the pBAD/Myc-His A plasmid (Invitrogen; for transcriptional control of the T7 RNA polymerase gene) and the GFPuv gene (Clontech; upstream of the T7 terminator region and downstream of the T7 promoter) between the *Ahd*II/*Bam*HI sites of plasmid pACYC177 (New England Biolabs). Plasmids pREP(1-12) were constructed by replacement of an *Hpa*I/*Apa*LI fragment of T7 RNA polymerase with overlap PCR fragments containing amber mutations at the positions described. Plasmid pREP/YC-JYCUA was constructed by ligation of an *Afe*I/*Sac*II fragment from pREP(10) and an *Ear*I(blunted)/*Sac*II fragment from pYC-J17 (Wang, L, Brock, A., Herberich, B. & Schultz, P.G., *Science*, 2001, 292, 498-500). The desired construct was identified following transformation into cells containing plasmid pQ screening for fluorescence.
- [302] Plasmid pQ was constructed by triple ligation of a *Aat*II/*Sal*I overlap PCR fragment containing the ScQRS downstream of the *lac* promoter region and upstream of the *E. coli* QRS termination region, a *Sal*I/*Ava*I overlap PCR fragment containing the *S. cerevisiae* tRNA(CUA)^{Gln} downstream of the *lpp* promoter region and upstream of an *rmbC* termination region, and the *Ava*I/*Aat*II fragment of pBR322 (New England Biolabs). Plasmid pQD was constructed by replacement of pQ fragment between *Bam*HI and *Bgl*II with a *Bam*HI/*Bgl*II fragment of the ScQRS (D291A) mutant.
- [303] Plasmid pBAD/TYAMB-4TAG was constructed by insertion of a PCR fragment of the S4Amber mutant of myoglobin, containing a C-terminal 6His-tag, into the pBAD/YC-JYCUA plasmid, a hybrid of plasmid pYC-J17 (Wang, L, Brock, A., Herberich, B. & Schultz, P.G., *Science*, 2001, 292, 498-500) and pBAD/Myc-His A (Invitrogen) containing the gene for *Mj*YtRNA_{CUA}, and the pBAD promoter and cloning regions for heterologous expression of an inserted gene.

WO 02/085923

PCT/US02/12465

[304] **Fluorimetric and cytometric analyses.** Single colonies containing desired plasmids were used to inoculate 2-mL GMML cultures containing the appropriate antibiotics, 0.002% Arabinose, and an appropriate unnatural amino acid, if desired. Cultures were grown to saturation and cells (200 μ L) were pelleted and resuspended in 1 mL phosphate-buffered saline (PBS). Cell concentrations were analyzed by absorbance at 600 nm and fluorescence levels were measured at 505 nm with excitation at 396 nm using a FluoroMax-2 fluorimeter. Cells suspended in PBS were analyzed cytometrically. To evaluate the permissivity of the amber positions within the T7 polymerase gene of pREP(10), the reporter plasmid was transformed into a panel of suppressor strains, which were subsequently analyzed fluorimetrically.

[305] **Evolution of aminoacyl-tRNA synthetase variants.** *M. jannaschii* TyrRS variants randomized at positions Y32, E107, D158, I159, and L162 (Wang, L, Brock, A., Herberich, B. & Schultz, P.G., *Science*, 2001, 292, 498-500) were transformed into DH10B *E. coli* cells (Life Technologies) containing pREP/YC-JYCUA to generate a library with a diversity of $\sim 10^9$. Transformants were allowed to recover in SOC medium for 60 min at 37°C, and were grown to saturation in LB medium. To begin an initial positive selection, 2 mL of library culture, pelleted and resuspended in GMML medium, was used to inoculate 500 mL of GMML containing 25 μ g/mL Tetracycline (Tet), 35 μ g/mL Kanamycin (Kn), and 1 mM pIF, pAF, pCF, or OAY. After incubation for 3 hr at 37°C, Cm was added to a final concentration of 75 μ g/mL and cells were grown to saturation (~ 48 hr). For the second positive selection, a 100-mL GMML culture containing Tet, Kn, 75 μ g/mL Cm, and 1 mM pIF, pAF, pCF, or OAY was inoculated with cells from the initial positive selection (500 μ L) and grown to saturation at 37°C (~ 24 –36 hr). In preparation for negative screening, a 25-mL GMML culture containing Tet, Kn, and 0.02% arabinose (Ara) was inoculated with cells from the second positive selection (100 μ L, pelleted and resuspended in GMML) and grown to saturation at 37°C (~ 24 hr). Ara-induced cells grown in the absence of unnatural amino acids (1 mL) were pelleted and resuspended in 3 mL of phosphate-buffered saline (PBS). Cells were sorted for lack of expression of GFPuv using a BDIS FACVantage TSO cell sorter with a Coherent Enterprise II ion laser with excitation at 351 nm and emissions detected using a 575/25 nm bandpass filter. Collected cells were diluted in at least 10 volumes of LB, containing Tet and Kn, and grown to saturation. To begin the third round of positive selection, 100 μ L of cells from the negative screen were pelleted, resuspended in GMML, and used to inoculate 25 mL of GMML containing Tet, Kn, and 1 mM pIF, pAF, pCF, or OAY. After

WO 02/085923

PCT/US02/12465

incubation for 3 hr at 37°C, Cm was added to a final concentration of 75 µg/mL and cells were grown to saturation (~24 hr). Following the third positive selection, cells were plated on GMML/agar containing Tet, Kn, 0.002% Ara, 0, 75, or 100 µg/mL Cm, and 0 or 1 mM pIF, pAF, pCF, or OAY, and grown for 48 hr at 37°C.

- 5 [306] **Expression and characterization of unnatural amino acid-containing proteins.** DH10B cells cotransformed with pBAD/JYAMB-4TAG and the appropriate pBK plasmid were used to inoculate a 100-mL GMML starter culture containing Kn and Tet, which was grown to saturation. A 500-mL culture containing Kn, Tet, 0.002% Ara, 5 µM FeCl₃, and the desired unnatural amino acid (or none) was inoculated with 50 mL of the starter
- 10 culture and grown to saturation (~18 hr). Cultures were pelleted, sonicated, and the myoglobin protein isolated according to the protocol of the QiaExpressionist (Qiagen) His-tag purification kit. Proteins were analyzed electrophoretically on a 12–20% gradient SDS polyacrylamide gel and by electrospray mass spectrometry.

Example 4: Creation of an autonomous 21 amino acid bacterium

- 15 [307] As described above, the common twenty amino acids are conserved across all known organisms. However, an expanded genetic code is provided herein, e.g., for added functionality, structure determination and the like. To determine whether the expanded genetic code is advantageous to a cell, e.g., with a particular unnatural amino acid, an autonomous bacterium that produces and incorporates the unnatural amino acid of interest
- 20 is desirable. The present invention provides such an autonomous twenty-one amino acid organism, and the results can be extended to the production of additional amino acid organisms, e.g., 22 amino acid organisms and the like. To produce an autonomous bacterium, three factors are typically considered: (i) the ability to synthesize a new amino acid from simple carbon sources; (ii) an aminoacyl synthetase that uniquely utilizes this
- 25 new amino acid and no other; and (iii) a tRNA that is acylated by that synthetase and no other, and which delivers the amino acid into proteins in response to a codon that does not encode any other amino acid.

- [308] A great deal of effort has been made toward in vivo incorporation of new amino acids to the genetic code but most of these do not have the incorporation specificity to
- 30 generate a healthy 21 amino acid bacterium. See, e.g., Hest, J.C.M.v., K.L. Kiick, and D.A. Tirrell, *J. Am. Chem. Soc.*, 2000. 122: p.1282; Hamano-Takaku, F., et al., *J. Biol. Chem.*, 2000. 275: p. 40324; and Budisa, N., et al., *FASEB J.*, 1999. 13: p. 41-51. However, it has recently been shown that that one could add new components to the

WO 02/085923

PCT/US02/12465

translational machinery of *E. coli* and site-specifically incorporate a variety of new amino acids into proteins *in vivo*, e.g., with high fidelity. See, e.g., Wang, L., et al., *Science*, 2001, 292: p. 498-500 and Wang, L. and P.G. Schultz, *Chem. Comm.*, 2002: p.1-10. See, also, co-filed patent application "Methods and Compositions for the Production of

5 Orthogonal tRNA-tRNA Synthetase Pairs," by Schultz et al. (Attorney Docket Number 54-000130), filed April 19, 2002.

[309] The present invention combines the above technology with a biosynthetic pathway system to produce an autonomous twenty-one amino acid bacterium. In addition, the present invention addresses the question of whether such organisms have or can be
10 evolved to have an evolutionary advantage over organisms that use the twenty natural amino acids.

[310] A completely autonomous bacterium typically comprises a biosynthetic pathway system, e.g., for producing an unnatural amino acid, and a translation system for incorporating the unnatural amino acid into one or more proteins in the bacterium. The
15 translation system typically comprises an aminoacyl synthetase that uniquely utilizes this unnatural amino acid and no other, and a tRNA that is acylated by that synthetase and no other, and which delivers the unnatural amino acid into proteins in response to a codon that does not encode any other amino acid. In one embodiment, the biosynthetic pathway system genes, aminoacyl synthetase genes, and tRNA genes are typically positioned on
20 separate plasmids to maximize control of the modified bacteria.

[311] In one example, the unnatural amino acid, *p*-aminophenylalanine (pAF), is biosynthetically produced and incorporated into proteins *in vivo*. pAF is optionally selected as a unnatural amino acid for an autonomous cell, e.g., based on its interesting physical properties, e.g., π donating effects, hydrogen bonding properties, and weak
25 basicity, its lack of toxicity to *E. coli*, and the fact that it is a known secondary metabolite. Moreover, the genes that lead to the production of pAF as a metabolic intermediate in the production of chloramphenicol and pristinamycin have been identified in *Streptomyces venezuelae* and *Streptomyces pristinaespiralis*, respectively. See, e.g., Yanai, K. and e. al., *Streptomyces venezuelae* genes *papA*, *papB*, *papC*, in PCT Int. Appl. 2001, Meiji
30 Seika Kaisha Ltd.: Japan. p. 1-83; and Blanc, V., et al., *Identification and analysis of genes from Streptomyces pristinaespiralis encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I*. Molecular Microbiology, 1997. 23(2): p. 191-202. As discussed above, pAF is optionally

WO 02/085923

PCT/US02/12465

synthesized in *E. coli* from chorismate 2 (a biosynthetic intermediate in the synthesis of aromatic amino acids) using the *S. Venezuelae* enzymes PapA, PapB, and PapC together with an *E. coli* aminotransferase. A plasmid, e.g., as provided in Figure 15A is optionally used to transform a cell to provide a cell that synthesizes its own supply of pAF *in vivo*.

- 5 An example plasmid for use in the biosynthesis of pAF *in vivo* is provided by SEQ. ID. NO.:67. SEQ ID NO.:68 provides the sequences for the individual genes papABC that encode the enzymes that are used to carry out the conversion of chorismate to pAF.
- [312] Once a cell is modified to produce an unnatural amino acid, e.g., pAF, O-methyl-L-tyrosine, a glycosylated amino acids, L-dopa or the like, the cell is also typically
- 10 modified by the addition of a translation system for incorporating the unnatural amino acid into one or more proteins within the cell. The translation system is typically provided to the cell via a separate plasmid than that by which the cell is modified to contain the biosynthetic pathway system as this allows closer control over the functions of the plasmids in the cell, e.g., regarding the number of copies, promoters, etc.
- 15 [313] The translation machinery typically comprises an orthogonal tRNA/RS pair, e.g., as provided by co-filed patent application "Methods and Compositions for the Production of Orthogonal tRNA-tRNA Synthetase Pairs," by Schultz et al. (Attorney Docket Number 54-000130), filed April 19, 2002. For example, an orthogonal tRNA/RS pair for pAF is optionally progenerated using a *Methanococcus jannaschii* tyrosyl-tRNA synthetase
- 20 (TyrRS) and mutant tyrosine amber suppressor tRNA (mtRNA^{Tyr}_{CUA}) pair as a starting point. See, e.g., Wang, L., et al., A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the *in vivo* incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.*, 2000 122: p. 5010-5011; and Wang, L. and P.G. Schultz, *Chem. and Biol.*, 2001, 8:883.
- 25 [314] For example, a pAF specific synthetase (pAFRS) is optionally generated by modifying the amino acid specificity of the *M. jannaschii* TyrRS to accept pAF and not any of the common twenty amino acids. See, e.g., Wang, L., et al., Expanding the genetic code of *Escherichia coli*. *Science*, 2001, 292: p. 498-500; Wang, L. and P.G. Schultz, Expanding the Genetic Code. *Chem. Comm.*, 2002: 1:1-10; and Wang, L., A. Brock, and
- 30 P.G. Schultz, Adding L-3-(2-naphthyl)alanine to the genetic code of *E. coli*. *J. Am. Chem. Soc.*, 2002. 124: p. 1836. A combination of positive selections and negative screens are optionally used to identify a pAFrs enzyme from a library of TyrRS variants containing random amino acids at five positions, e.g., Tyr32, Glu107, Asp158, Ile159, and Leu162.

WO 02/085923

PCT/US02/12465

- A single reporter plasmid is optionally used for both selection and screening, e.g., as described in co-filed patent application "Methods and Compositions for the Production of Orthogonal tRNA-tRNA Synthetase Pairs," by Schultz et al. (Attorney Docket Number 54-000130), filed April 19, 2002. The positive selection is typically based on suppression of a TAG codon at a permissive position within the chloramphenicol acetyltransferase (CAT) gene. (see, e.g., Wang, L., et al., Expanding the genetic code of *Escherichia coli*. *Science*, 2001, 292: p. 498-500 and Pasternak, M., T.J. Magliery, and P.G. Schultz, A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code. *Helvetica Chimica Acta*, 2000 83: p. 2277), e.g., by either pAF for an endogenous amino acid. Cells containing the TyrRS library and reporter plasmid grown in liquid culture containing pAF are typically selected for survival, e.g., in the presence of chloramphenicol (Cm). The negative screen based on suppression of two UAG stop codons at permissive positions within the T7 RNA polymerase gene drives the expression of green fluorescent protein (GFP). Positively selected cells grown in the absence of pAF and Cm, are then typically screened, e.g., using fluorescence activated cell sorting (FACS) for the lack of fluorescence.
- [315] *Evolution of pAFrs*: The reporter plasmid, pREP(2)/YC-JYCUA, contains the genes for CAT, T7 RNA polymerase, GFP, and mtRNA^{Tyr}_{CUA}, and a selectable marker for Tet resistance (Santoro unpublished results). The CAT gene contains a TAG codon substitution at position D112. The T7 RNA polymerase gene contains a seven-amino acid N-terminal leader peptide and TAG substitutions at M1 and Q107. For the positive selection, cells were grown in GMML minimal media containing 35 µg/ml Kn, 25 µg/ml Tet, 75 µg/ml Cm, and 1mM pAF (Sigma). For the negative screen, cells were grown in GMML media containing 35 µg/ml Kn, 25 µg/ml Tet, and 0.002 % arabinose. FACS was carried out using a BDIS FACVantage TSO cell sorter with a Coherent Enterprise II ion laser. The excitation wavelength was 351 nm and emission was detected using a 575/25 nm bandpass filter. Collected cells were diluted into at least 10 volumes of LB, containing Tet and Kn, and grown to saturation.
- [316] *Addition of pAF biosynthetic pathway*: The papA, papB, and papC genes were PCR amplified from *S. venezuelae* (ATCC 10712) genomic DNA. Genes, papABC were assembled by overlap PCR and inserted into a pSC101 derived plasmid, pLASC, and maintained by ampicillin selection. Ribosome binding sites (rbs) were from the 5' UTR of LacZ, malE, and cro and placed prior to papA, papB, and papC, respectively. The

WO 02/085923

PCT/US02/12465

papABC genes were placed under control of lac and lpp promotor to afford two pathway plasmids pLASC-lacPW and pLASC-lppPW.

[317] *Testing pAF biosynthesis with pAFRS*: E. coli DH10B cells harboring three plasmids, the reporter plasmid (pREP(2)/YC-JYC UA), the synthetase (pAFRS), and the pathway plasmid (pLASC-lacPW or pLASC-lppPW) were grown to saturation in GMML minimal media (pLASC was used for background, no pAF, and 1 mM exogenous pAF trials). DH10B was grown with no plasmids to determine the background suppression level of the reporter plasmid. A sample of each cell growth was diluted to an OD of 1.0 (600 nm) with water and 200 μ L was pelleted. Cell were suspended in 1 mL 1% PBS and analyzed using a Fluoromax-2 fluorescent detector (excitation wavelength was 351 nm and a peak emission at 505 nm was monitored). DH10B produced 1.0×10^4 fluorescent units, while background fluorescence (no pAF added) from the reporter system produced 2.5×10^4 fluorescent units. The lacPW, lppPW, and 1 mM exogenously added pAF produced 7.9×10^4 , 3.0×10^6 , and 3.0×10^4 fluorescent units, respectively. Induction of the lacPW with IPTG was not feasible due its inhibitory affect on the arabinose promotor in the reporter plasmid, (pREP(2)/YC-JYC UA).

[318] *Aromatic amino acid concentration*: E. coli DH10B cells harboring the pLASC plasmid and pLASC-lacPW or pLASC-lppPW were grown in GMML minimal media (1% glycerol, 0.3mM leucine) containing 110 μ g/ml ampicillin to saturation. Cells grown with exogenously added pAF contained 1 mM amino acid at the start of the growth. Cells were harvested by centrifugation (100 ml), washed, 1 ml of water and 0.2 ml of toluene was added. Cells were shaken at 37 °C 11for 30 minutes and then separated by centrifugation. The aqueous layer was filtered (microcon YM-10) and analyzed by HPLC-MS (Agilent 1100): 5-15 μ L of the aqueous layer seperated on Zorbax SB-C18 column (5 μ m, 4.6X150mm) with a gradient of water 1% TFA /acetonitrile 1% TFA (95:5) to (5:95) over 10 minutes. Amino acids were identified by abstracting their MW(+1) from the total ion mass spectrum. The area of the abstracted ion was used to calculate amount of amino acids present in each sample. Cellular concentrations were based on the amount of water in the cell pellet, 70% by mass.

[319] *Expression of protein containing pAF*: Plasmid pBAD/JYAMB-4TAG with tetracycline resistance was used to express the Tyr CUA mutRNA gene under the control of the lpp promotor and rrnC terminator, and the myoglobin gene (with an amber stop codon at Ser4) under the control of the arabinose promotor and rrnB terminator. A his6-tag was added to the carboxy terminus of myoglobin. The TyrRS and pAFRS genes were

WO 02/085923

PCT/US02/12465

expressed under the control of the *E. coli* GlnRS promoter and terminator on a pBR322 derived plasmid with kanamycin resistance. The papABC genes were expressed from pLASC-lacPW or pLASC-lppPW (13) under the control of the native terminator. *E. coli* DH10B cells harboring plasmid pBAD/JYAMB-4TAG, pBK-TyrRS or pBK-pAFRS, and a pLASC derived plasmid (pLASC, pLASC-lacPW or pLASC-lppPW as indicated) were grown in 0.5 L of minimal media containing 0.002 % arabinose. Expression trials with exogenous pAF contained a final concentration of 1 mM pAF (Sigma). For all trials, cells were grown to saturation (20-30 hrs) in parallel at 37 °C, pelleted, and protein was purified by Ni +2 affinity chromatography according to manufacturer's protocol under native conditions (Qiagen, Valencia, Ca). Fifteen µl of final protein solution (3.5 ml) from each preparation were separated on a 12% SDS polyacrylamide gel and silver-stained.

Example 5: In vivo incorporation of O-methyl-L-tyrosine in an *E. coli* cell which has been genetically engineered to biosynthesize the unnatural amino acid

[320] As discussed herein, one aspect of the invention is biosynthetic pathways for unnatural amino acids in *E. coli*. This is accomplished by e.g., addition to the cell of genes for new enzymes or modification of existing *E. coli* pathways. In this example, *E. coli* was genetically engineered to produce the unnatural amino acid O-methyl-L-tyrosine.

[321] Plant O-methyltransferases are enzymes involved in secondary metabolism, which converts a hydroxyl group into a methoxyl group. Two enzymes, (iso)eugenol O-methyltransferase (IEMT) and caffeic acid O-methyltransferase (COMT) (Clarkia brewery) were selected for incorporation into *E. Coli*. IEMT methylates eugenol/isoegenol, and COMT methylates caffeic acid. The substrates of these two enzymes are similar to tyrosine. However, both enzymes have high substrate specificity and methylation regiospecificity.

[322] A combinatorial approach was used to evolve the substrate specificity of both enzymes to tyrosine, thereby converting tyrosine to O-methyl-L-tyrosine. Active sites of the proteins were mutated to produce large mutant libraries and several rounds of selection were completed. Three clones were identified. The clones are characterized and at least one is selected to generate an *E. coli* strain that biosynthesizes O-methyl-L-tyrosine. This strain of *E. coli* is genetically engineered to also express the orthogonal tRNA/RS pair described in Example 1 above, thereby providing a cell for autonomous in vivo incorporation of an unnatural amino acid.

WO 02/085923

PCT/US02/12465

Example 6: In vivo incorporation of heavy atom amino acids

- [323] Structure-guided drug discovery has historically been a slow, laborious process used in only a modest fraction of drug discovery programs in the industry. One bottleneck is the phase problem encountered when using X-ray crystallography to solve protein
- 5 structure. Typically, the protein has to be expressed again in the presence of selenomethionine, which doubles the work load and may not necessary result in successful crystallization. An alternative method is to soak the crystal in a heavy-atom-containing solution, which may result in crystal crush. In vivo incorporation of heavy-atom containing unnatural amino acids into proteins is a useful tool to accelerate the solving of
- 10 protein crystal structures.
- [324] The site specific in vivo incorporation of p-iodo-phenylalanine and p-bromo-phenylalanine into proteins was performed. Iodine and bromine are heavy atoms, and the incorporation facilitates solving of phase using MAD. The site-specific introduction of heavy atoms using unnatural amino acids also provides selectivity and flexibility in
- 15 choosing positions for heavy atoms.
- [325] Mutant synthetases with specificities for p-iodo-phenylalanine and p-bromo-phenylalanine, respectively, were generated following the methods and compositions described in Example 1. The protein Z domain (B. Nilsson, et al, *Protein Eng.* 1:107-113 (1987)) was expressed, in which bromine or iodine was selectively introduced in the form
- 20 of p-iodo-phenylalanine and p-bromo-phenylalanine using in vivo incorporation of the unnatural amino acids. Protein crystal trays were set up following standard protocols.
- [326] The three dimensional structure of the protein is solved using X-ray crystallography; the phase is determined using the heavy atoms present in the protein.

Example 7: In vivo incorporation of meta-tyrosine analogues

- 25 [327] An orthogonal TyrRS was generated for aminoacylation of the mtRNA_{CUA}^{Tyr} (described in Example 1) with meta-tyrosine analogues.
- [328] **Preparation of mutant TyrRS library plasmids.** A library of plasmids encoding mutant *M. jannaschii* TryRSs directed at meta-substituted tyrosine derivatives was constructed, generally following the methods described in Example 1. Briefly, six
- 30 residues (Tyr³², Ala⁶⁷, His⁷⁰, Gln¹⁵⁵, Asp¹⁵⁸, Ala¹⁶⁷) in the active site of *M. jannaschii* TyrRS that are within 6.9 Å of the meta-position of the aryl ring of bound tyrosine in the crystal structure of *Bacillus stearothermophilus* TyrRS were mutated to all 20 amino acids

WO 02/085923

PCT/US02/12465

at DNA level using the NNK codon scheme as described in Example 1 above. The constructed plasmid library pBK-lib contained around 1×10^9 independent clones.

[329] Evolution of orthogonal tRNA-synthetase pairs for incorporation of *m*-acetyl phenylalanine. After 3 rounds of positive selection and 2 rounds of negative selection,

5 five candidate clones (SEQ ID NO: 17-21) emerged whose survival in chloramphenicol was dependent on the addition of the unnatural amino acid. In the absence of *m*-acetyl phenylalanine, the IC_{50} of chloramphenicol resistance for cells harboring the one of the three mutant TyrRS plasmids is 20 $\mu\text{g/ml}$. In the presence of *m*-acetyl phenylalanine, the IC_{50} of resistance to chloramphenicol for the same cells is 100 $\mu\text{g/ml}$. The large difference
10 between these two numbers reflects the ability of the selected synthetases to specify the incorporation of *m*-acetyl phenylalanine over the natural amino acids in the cell. The data for *m*-methoxy phenylalanine were similar; five clones were isolated (SEQ ID NO:22-26).

[330] Protein expression of unnatural amino acid incorporated DHFR. The *m*-methoxy phenylalanine and *m*-acetyl phenylalanine synthetases selected above were used
15 to incorporate the relevant unnatural amino acids in response to an amber codon in DHFR as previously described in Example 1 above. As a negative control, cells containing both the orthogonal pair of tRNA-synthetase and amber-mutational vector encoding DHFR were grown in the absence of unnatural amino acids. The results of protein expression are shown in Figure 10. These results clearly demonstrated the specificity of the orthogonal
20 pair of tRNA-synthetase to incorporate unnatural *m*-methoxy phenylalanine and *m*-acetyl phenylalanine. The yields of expressed DHFR protein are approximately 0.5 mg/L of culture in both cases.

[331] Utilizing *meta*-acetyl phenylalanine as a chemical handle. The *m*-acetyl phenylalanine incorporated DHFR protein was labeled with hydrazide derivatives, both
25 extra-cellularly and intra-cellularly at a milligram scale. The carbonyl group will react rapidly with hydrazide in aqueous solution to form hydrazone that is stable under physiological conditions (Shao, J.; Tam, J. J. Am. Chem. Soc. 117, 3893-3899 (1995)). This chemistry has been used by Schultz and coworkers to specifically label a ketone containing, purified T4 lysozyme with fluorescein hydrazide (Cornish, V. W.; Hahn, K.
30 M.; Schultz, P. G. J. Am. Chem. Soc. 118, 8150-8151 (1996)).

[332] Purified *m*-acetyl phenylalanine-incorporated DHFR protein was treated with fluorescein hydrazide in aqueous buffer. As a control in parallel, a purified *m*-methoxy phenylalanine-incorporated DHFR protein was subjected to the same reaction conditions.

WO 02/085923

PCT/US02/12465

After the reaction, both proteins were purified and then excited at 491 nm to obtain fluorescence emission spectra shown in Figure 11. Under identical conditions, the purified *m*-acetyl phenylalanine-incorporated DHFR was labeled with fluorescein hydrazide while *m*-methoxy phenylalanine was not labeled.

- 5 [333] The fluorescein hydrazide is cell-permeable and does not lyse cells at 4 °C. Thus, it is possible to label the *m*-acetyl phenylalanine-incorporated DHFR protein intracellularly with fluorescein hydrazide. Cells expressing the "ketone handle"-incorporated DHFR were incubated with fluorescein hydrazide solution. After 36 hours at 4 °C and extensive washes to remove excess fluorescein hydrazide, the labeled DHFR protein was
- 10 purified and subjected to fluorescence emission tests. As a negative control in parallel, *m*-methoxy phenylalanine-incorporated DHFR was also purified with the same procedures. Similar results to the extracellular experiment (Figure 15) were obtained when intact cells were labeled with fluorescein hydrazide and the DHFRs were subsequently purified.
- [334] These experiments demonstrated one example of the utility of a protein with at least one unnatural amino acid. Other compounds can be used to *in vivo* label proteins
- 15 with at least one unnatural amino acid. Examples include, e.g., biotin hydrazide and other hydrazide derivatives.

Example 8: In vivo incorporation of photoreactive amino acids

- [335] **Introduction:** Experiments were performed in which photocrosslinker amino
- 20 acids were genetically encoded and site specifically incorporated into a specific protein *in vivo*. This protein was then crosslinked at will by excitation of the photoreactive group-providing temporal control.
- [336] This invention is useful for, e.g., exploring protein interactions. For example, this invention is useful for defining residues in the protein primary sequence that mediate
- 25 interaction with different cellular components by varying the position of the crosslinker in the protein. Because a covalent bond is formed between the protein and the molecule it interacts with it is possible to detect weak or transient interactions.
- [337] Two chemical functional groups have gained prominence as crosslinkers, aryl-azides and benzophenones since they can be activated at wavelengths above 300 nm (
- 30 below which protein damage via photooxidation may be a problem). These two crosslinking groups were been incorporated into the unnatural amino acids *p*-azido-phenylalanine and *p*-benzoyl-phenylalanine respectively (Figure 12).

WO 02/085923

PCT/US02/12465

- [338] **Generation of O-RS specific for photocrosslinker amino acids.** The orthogonal pair described in Example 1, *Methanococcus jannaschii* mtRNA^{Tyr}_{CUA}/TyrRS pair was used as the starting point to generate an O-RS specific for the crosslinker-unnatural amino acid p-azido-phenylalanine (pBpa). The methods for mutagenesis, screening and selection were performed following the experimental outline described in Example 1. Briefly, a MjTyrRS library of mutants was generated in which five residues (Tyr 34, Glu 107, Asp 158, Ile 159, Leu 162) were randomized. These residues were chosen on the basis of the crystal structure of *Bacillus Stearotherophilus* TyrRS complexed with tyrosyl adenylate (P.Brick, T.N.Bhat & D.M.Blow *Journal of Molecular Biology* 208, 83 (1989)) in which homologous residues (Tyr34, Asn123, Asp176, Phe177, Leu180) are within 6 Å of the para position of the aryl ring of bound tyrosine. The mutant TyrRS library was passed through a positive selection based on suppression of an amber stop codon at a permissive site (Asp112) in the chloramphenicol acetyl transferase (CAT) gene. Cells transformed with the synthetase library, and the CAT mutant were challenged to grow in the presence of 1mM *pBpa* and chloramphenicol. Surviving cells contained synthetases capable of charging the orthogonal mtRNA^{Tyr}_{CUA} with either a natural or unnatural amino acid. These synthetase genes were transferred into cells containing mtRNA^{Tyr}_{CUA} and a variant of the gene encoding the toxic barnase protein, which contains three amber mutations at permissive sites (Gln2, Asp44, Gly65)(Wang, L., Brock, A., Herberich, B. & Schultz, P. G. *Science* 292, 498-500(2001)). Growth of these cells in the absence of *pBpa* selected against synthetases capable of utilizing natural amino acids.
- [339] After five rounds of positive and negative selection the surviving synthetase plasmids were transformed into a reporter strain in which the production of full length CAT and T7 RNA polymerase (T7 RNAP) are dependent on suppression of amber stop codons in the CAT and T7 RNAP gene, respectively (Santoro SW, Schultz PG. *Proc Natl Acad Sci U S A* Apr 2;99(7):4185-90 (2002)). Because the T7 RNAP drives expression of the green fluorescent protein (GFP) these cells can be fluorometrically screened. Ninety-six clones were screened for pBpa dependent chloramphenicol resistance and GFP fluorescence. Six distinct synthetases conferred Ile chloramphenicol resistance on *E.coli* with IC₅₀s of 120 mg/L and 5 mg/L in the presence and absence of 1mM *pBpa* respectively; they also showed *pBpa* dependent GFP fluorescence. The large difference between the chloramphenicol resistance in the presence and absence of *pBpa* shows a

WO 02/085923

PCT/US02/12465

substantial *in vivo* specificity of the selected synthetase/tRNA pairs for insertion of *pBpa* over all twenty natural amino acids found in the cell in response to an amber codon.

- [340] **In vivo incorporation of *pBpa* into myoglobin.** To measure the fidelity and efficiency of *pBpa* incorporation, the codon for Ser4 in sperm whale myoglobin (containing a C-terminal His6 tag) was converted to an amber codon. In the presence of
- 5 both Mj p-BpaRS-1, mtRNA_{CUA}^{Tyr} and *pBpa*, full length myoglobin was produced with a purified yield of 2 mg/L. No myoglobin protein was detectable by silver stain or Western blot against the C-terminal His6 tag on myoglobin if any of the three components responsible for specific amber suppression with *pBpa* (amino acid, synthetase, or tRNA)
- 10 were withheld. This data provides further evidence that the selected synthetase is very selective for *pBpa*.

- [341] Electrospray-ionization ion trap mass spectrometry of the mutant myoglobin gave a mass of 18519 ± 0.5 which is identical to the calculated mass of 18519.0 for the *pBpa* containing protein. This confirms the incorporation of *pBpa* at a single site in the protein.
- 15 No masses were observed in the mass spectra corresponding to natural amino acid incorporation providing additional evidence for the high fidelity incorporation of *pBp*.

- [342] **Sequence analysis of mutant O-RS.** The selected synthetases show interesting sequence convergence. Tyr32 of *M. jannaschii* TyrRS is converted to alanine or glycine in five of the six mutant synthetase clones. Asp158 of the *M. jannaschii* TyrRS is
- 20 converted to threonine in five of the six selected mutants, while Ile159 is converted to serine in four of the six mutants. Serine or proline substitutions dominate at position 107 of *M. jannaschii* TyrRS; Leu162 is conserved in four of the six mutants. A consensus set of mutations (32:Gly, Ala/ 107:Ser, Pro/ 158:Thr/ 159: Ser/ 162: Leu) emerges from this analysis.

- [343] **In vivo incorporation of *pBpa* into GST.** To demonstrate the utility of this methodology for mapping protein-protein interactions, a crosslinking experiment was carried out with glutathione-S-transferase. This protein is a dimer of two identical subunits which have previously been crosslinked non-specifically using glutaraldehyde. The crystal structure of the dimeric *Schistosoma Japonica* glutathione-S-transferase
- 30 (SjGST) (McTigue, M. A., Williams, D. R. & Tainer, J. A. *Journal of Molecular Biology* 246, 21-27(1995)) was used to identify two sites to substitute with *pBp*: residue Phe52, which is buried in the dimer interface of the crystal structure, and residue Tyr198 which is solvent exposed. The codons corresponding to Phe52 or Tyr198 in the gene for a 27 kDa

WO 02/085923

PCT/US02/12465

protein Sj GST, were replaced with amber codons. The orthogonal synthetase tRNA pair was then used to site specifically incorporate *pBpa* into SjGST in *E.coli* at these sites.

Upon irradiation with long wavelength ultraviolet radiation, purified SjGST was converted to a covalently linked homodimer as judged by denaturing SDS PAGE. Approximately 70

- 5 % of the SjGST present was crosslinked in 5 minutes. In contrast, control experiments using either wild type SjGST or SjGST containing *pBpa* at residue 198, which lies outside the dimer interface, shows no detectable crosslinking in response to UV irradiation.
- [344] These results demonstrate that site-specific *pBpa* substitution can be used to define amino acids involved in a protein-protein interaction.

- 10 [345] **Characterization of Mutant Synthetases** Individual synthetase clones in DH10B/ pREP(2)/YC-JYCUA were used to inoculate 0.5 mL of LB supplemented with kanamycin and tetracycline to 30, 20 mg/L. After 20 hours growth (37°C, 300 rpm) cells were diluted 10⁴ fold in dh20 and replica spotted on two sets of GMML plates. One set of
- 15 plates were supplemented with kanamycin and tetracycline at 30 and 20 micrograms/L, respectively, and chloramphenicol at concentrations ranging from 0 micrograms/L to 110 micrograms/L. The second set of plates were identical to the first, except that they were supplemented with 1 mM *pBpa*. After 48 h the I₅₀ of chloramphenicol resistance in the presence and absence of *pBpa* was calculated from the concentration of chloramphenicol at which half the number of colonies on the plates with no chloramphenicol were visible.
- 20 GFP expression in the presence and absence of *pBpa* was imaged using a Storm phosphorimager(Molecular dynamics). Mutant synthetase genes exhibiting the strongest amino acid dependence in both GFP signal and chloramphenicol resistance were isolated and sequenced by standard methods.

- [346] **Protein Expression** Plasmid PYC/SjGSTmut, which contains the mutant SjGST
- 25 gene on an arabinose promoter and *rrnB* terminator, and *mtRNA*_{CUA}^{Trp} on a *lpp* promoter and *rrnC* terminator, and a tetracycline resistance marker was co-transformed with a pBK vector expressing p-BpaRS into DH10B *E.coli*. Cells were amplified in 10 mL of 2x-YT containing kanamycin at 30 micrograms/L and tetracycline at 25 micrograms/L before being washed in PBS and used to inoculate 1 L of liquid GMML with the appropriate
- 30 antibiotics and *pBpa* to 1 mM. Protein expression was induced at an O₆₀₀ of 0.6 by the addition of arabinose to 0.2% followed by 5 hours growth. Cells were harvested by centrifugation and protein was purified by virtue of a C-terminal hexa-histidine tag using Ni-NTA affinity chromatography.

WO 02/085923

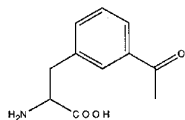
PCT/US02/12465

- [347] Sperm whale myoglobin was expressed and purified from cells containing pBAD/JYAMB-4TAG in an analogous manner to SjGST, except that induction was constitutive with 0.002 % arabinose. Samples for mass spectrometry were desalted on a NAP-10 column (Pharmacia) and purified by HPLC. To verify the incorporation of *pBpa*, the protein mass was ascertained by electrospray-ionization ion trap mass spectrometry.
- [348] **Mutant Sj GST Cloning** Mutant SjGST genes were assembled by overlapping PCR, using pGEX-3 (Pharmacia) as a template. All PCR reactions were carried out using the Expand PCR kit (Roche) according to the manufacturers instructions. The resulting genes were digested with Nco I and Kpn I restriction enzymes and cloned into predigested, dephosphorylated pBADJYC vector between the same restriction sites and in frame with a C-terminal hexa-histidine tag. All final constructs were confirmed by DNA sequencing.
- [349] **Photo-activated crosslinking.** Crosslinking reactions were performed in a 96 well microtitre plate (Nuncosorb) using 100 μ L of 10 ng/ μ L SjGST (in 50 mM NaH₂PO₄, 300 mM NaCl, 250mM imidazole) at 4 °C. Samples were irradiated at 365 nm using a handheld UV lamp (115V, 60 Hz, 0.2 A; Spectronics, NY, USA), for 1 min or 5 min. Samples were removed from the wells and diluted with SDS loading buffer before resolution of products by SDS-PAGE on a 10-20 % gradient gel. SjGST was transferred to PVDF (Biorad) and probed by western blot using goat anti-GST (Pharmacia) and a secondary mouse anti goat HRP conjugate (Sigma). Signal was developed using Super signal West (Pierce) and visualized by exposure on hyperfilm (Amersham).

Example 9: Synthesis of meta-substituted phenylalanines

[350] In one aspect, the present invention provides meta substituted phenylalanines as shown in Formula IV:

IV

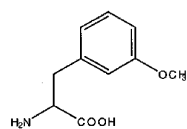


- [351]
- [352] and in Formula V.:

WO 02/085923

PCT/US02/12465

V



[353]

[354] Formula IV illustrates the structure of 3-acetyl-phenylalanine and Formula V represents 3-methoxy-phenylalanine.

[355] Meta-substituted phenylalanines are synthesized in a procedure as outlined in

- 5 Figure 14. Typically, NBS (N-bromosuccinimide) is added to a meta-substituted methylbenzene compound to give a meta-substituted benzyl bromide, which is then reacted with a malonate compound to give the meta substituted phenylalanine. Typical substituents used for the meta position include, but are not limited to, ketones, methoxy groups, alkyls, acetyls, and the like. A specific example is provided below.

- 10 [356] NBS (N-bromosuccinimide) was recrystallized from boiling water prior to usage. NBS (1.85 g, 10.5 mmol) was added to a solution of 3-methyl acetophenone (1.34 g, 10 mmol). AIBN (2', 2'-azobisisobutyronitrile) (0.043 g, 0.25 mmol) was added to the mixture. The reaction mixture was refluxed for 4 hours. The completion of reaction was checked by TLC (8:1/hexanes: EtOAc). After aqueous workup, the organic solvent was removed and hexanes was added to give solid. The solid was filtered and washed with
- 15 hexanes and EtOAc. Then the mixture was recrystallized with hexanes. The supernatant was collected and solvent was removed to give compound (1-(3-bromomethyl-phenyl)-ethanone).

- [357] Dry ethanol (50 ml) was added dropwise to pentane-washed sodium pieces (2.3 g, 0.1 mol) under argon atmosphere. After the completion of addition, stirring was required to dissolve the last pieces of sodium. A solution of diethyl acetylamido-malonate ester (21.7 g, 0.1 mol) was added over 30 minutes. 1-(3-bromoethyl-phenyl)ethanone (21.1 g, 0.1 mol) in dry ethanol was added dropwise over 90 minutes. After the mixture was refluxed overnight, ether and water was added, and the organic layer was separated. After
- 20 aqueous workup, the organic layers were combined, washed with brine, dried over MgSO_4 and filtered. The solvents were removed in vacuo. Hexanes-dichloromethane, 4:1, was added to the residue, and the insoluble material was filtered out and washed exhaustively with 10:1 dichloromethane-benzene to give diethyl 2-acetamido-2[3-acetyl-phenyl]-
- 25

WO 02/085923

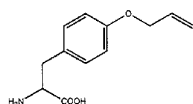
PCT/US02/12465

- methyl]malonate. This compound was stirred with 8 M HCl in dioxane overnight. Then the mixture was taken to dryness, water was added, and it was taken to dryness again to give final compound m-acetylphenylalanine hydrochloride. HPLC was used to purify the desired compound as white solid. The total yield was 64%. ¹HNMR (D₂O): δ 7.85-7.28 (m, 4H), 4.23 (dd, 1H), 3.2 (m, 2H), 2.7 (s, 3H). Calculated molecular weight: 243.69, obtained molecular weight: 243.07. A similar synthesis is used to produce a 3-methoxy phenylalanine. The R group on the meta position of the benzyl bromide in that case is –OCH₃. See, e.g., Matsoukas et al., *J. Med. Chem.*, 1995, 38, 4660-4669.

Example 10: Synthesis of 4-allyl-L-tyrosine

- [358] In another aspect, the present invention provides 4-allyl-L-tyrosine, whose structure is shown in Formula II:

II



[359]

- [360] The compound of Formula II, 4-allyl-L-tyrosine, is synthesized according to the scheme set forth in Figure 13. A protected tyrosine, e.g., an N-boc or Fmoc protected tyrosine, is reacted with allyl bromide, resulting in a protected allyl tyrosine, which is then typically deprotected to yield 4-allyl-L-tyrosine. For example, N-(tert-Butoxycarbonyl)-L-tyrosine (2.95g, 10 mmole) was dissolved in 80 ml of DMF. The solution was chilled to 5 °C and NaH (0.63g, 26 mmole) was added. The reaction mixture was allowed to warm up to 10 °C and stirred for additional 2 hours. After that, allyl bromide (1.33g, 11 mmole) was added to the mixture and reaction was warmed to room temperature. The reaction mixture was stirred for 4 hours. Water was added to work up the reaction. The aqueous layer was extracted with ethyl acetate and CH₂Cl₂. The organic layer was dried over anhydrous MgSO₄. The organic solvent was removed to give white solid. This compound was then refluxed in 4M HCl in 1,4-dioxane for 4 hours. All the solvent was evaporated to give the desired product as white solid (1.9g, 86%). ¹HNMR (CD₃OD): δ ppm 3.1(m, 2H), 4.1 (t, 1H), 4.5 (d, 2H), 5.3 (q, 1H), 5.9 (m, 1H), 6.9 (d, 2H), 7.1 (d, 2H). Calculated molecular weight: 221, obtained molecular weight: 222.

WO 02/085923

PCT/US02/12465

Example 11: Cellular uptake screen of unnatural amino acids

- [361] A variety of unnatural amino acids and α -hydroxy acids of interest, obtained commercially or by short syntheses from available starting materials (I. Shin, B. Herberich, A. Varvak, T. Magliery, P. Schultz, unpublished results, were screened for cell toxicity). For example, Figure 29 provides a library of unnatural amino acids useful for the following screen. Each amino acid was screened at 1 mM in glycerol minimal media for toxicity to cells, e.g., to DH10B harboring pBLAM-YQRS and pACYsupA38. Toxicities are sorted into five groups: (1) no toxicity, in which no significant change in doubling times occurs; (2) low toxicity, in which doubling times increase by less than about 10% (seen with the following compounds in Figure 29: S63, S69, S74, S75, S81, S95); (3) moderate toxicity, in which doubling times increase by about 10% to about 50% (seen in the following compounds shown in Figure 29: B, M, P, S12, S14, S22, S41, S45, S49, S52, S62, S64, S65, S71, S91, S93, B10); (4) high toxicity, in which doubling times increase by about 50% to about 100% (seen in the following compounds from Figure 29: C, Q, V, BB, S2, S5, S50, S60, S78, S83, S89, S90); and (5) extreme toxicity, in which doubling times increase by more than about 100% (observed for the following compounds from Figure 29: W, S15, S26, S27, S30, S31, S39, S47, S88, S94). See, e.g., Liu, D.R. & Schultz, P.G. Progress toward the evolution of an organism with an expanded genetic code. *Proceedings of the National Academy of Sciences of the United States of America* 96, 4780-4785 (1999).
- [362] The toxicity of amino acids scoring as highly or extremely toxic is typically measured as a function of their concentration to obtain IC₅₀ values. In general, amino acids which are very close analogs of natural amino acids (e.g., Q, W, S5, S26, S27, S50, S90, S94) or which display reactive functionality (e.g., S15, S39, S47) demonstrated the highest toxicities.
- [363] To identify possible uptake pathways for toxic amino acids, toxicity assays were repeated at IC₅₀ levels (typically 3 μ M to 500 μ M) in media supplemented with an excess (2 mM) of a structurally similar natural amino acid. For toxic amino acids, the presence of excess natural amino acid rescued the ability of the cells to grow in the presence of the toxin, presumably because the natural amino acid effectively outcompeted the toxin for either cellular uptake or for binding to essential enzymes. In these cases, the toxic amino acid can be assigned a possible uptake pathway and labeled a "lethal allele" whose complementation is required for cell survival. Lethal alleles identified in this manner (16

WO 02/085923

PCT/US02/12465

of the toxic unnatural amino acids) span ten possible amino acid uptake groups: alanine, glutamic acid, lysine, leucine, methionine, proline, glutamine, arginine, threonine, and tyrosine.

- 5 [364] These lethal alleles are extremely useful for assaying the ability of cells to uptake nontoxic unnatural amino acids. Each nontoxic unnatural amino acid was added at 2 mM to media containing IC50 levels of each lethal allele. Complementation of the toxic allele, evidenced by the restoration of cell growth, shows that the nontoxic amino acid is taken up by the cell, possibly by the same uptake pathway as that assigned to the lethal allele. A lack of complementation is inconclusive.
- 10 [365] Using this method, the ability of 22 glutamine and glutamic acid analogs to be taken up by DH10B was evaluated. Amino acids S27 and S47 were used as toxic glutamine alleles at 100 μ M and 30 μ M, respectively, while S50 was employed as a toxic glutamic acid allele at 150 μ M. Results from S27 and S47 complementation were in complete agreement and identified amino acids B, Z, S6, S60, S61, and S62 (in addition to
- 15 S27 and S47) as being uptaken by cells possibly via the glutamine uptake pathway. Similarly, complementation of S50 identified B, C, K, X, S60, S65, and S84 as being uptaken into DH10B, possibly via the glutamic acid transport system.
- [366] These findings indicate that the *E. coli* glutamine and glutamic acid transport pathways may tolerate significant perturbations in amino acid structure, including side
- 20 chain elongation (X and Z), ketone or methylene placement at the γ -position (B, C, S65), carboxamide replacement with a sulfoxide (S61), a known substrate for a bacterial glutamine transporter or hydrazide (S47), also a known glutamine transporter substrate as well as a variety of hybridization changes at the side chain terminus (S60, S62, K, S84). See, e.g., Jucovic, M. & Hartley, R.W. Protein-protein interaction: a genetic selection for
- 25 compensating mutations at the barnase-barstar interface. *Proceedings of the National Academy of Sciences of the United States of America* 93, 2343-2347 (1996) and Weiner, J.H., Furlong, C.E. & Heppel, L.A. A binding protein for L-glutamine and its relation to active transport in *E. coli*. *Archives of Biochemistry and Biophysics* 142, 715-7 (1971).

Example 12: Biosynthesis of *p*-aminophenylalanine

- 30 [367] To produce the unnatural amino acid *p*-aminophenylalanine (pAF) in vivo, genes relied on in the pathways leading to chloramphenicol and pristinamycin are optionally used. For example, in *Streptomyces Venezuelae* and *Streptomyces pristinaespiralis*, these genes produce pAF as a metabolic intermediate. See, e.g., Yanai, K. and e. al.,

WO 02/085923

PCT/US02/12465

Streptomyces venezuelae genes *papA*, *papB*, *papC*, in PCT Int. Appl. 2001, Meiji Seika Kaisha Ltd.: Japan, p. 1-83; and Blanc, V., et al., *Identification and analysis of genes from Streptomyces pristinaespiralis encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I*. Molecular Microbiology, 1997, 23(2): p. 191-202.

[368] A biosynthetic pathway for pAF is shown in Figure 15, Panel B. pAF is optionally synthesized in *E. coli* from chorismate (compound 2 In Figure 15, Panel B), which is a biosynthetic intermediate in the synthesis of aromatic amino acids. To synthesize pAF from chorismate, a cell typically uses a chorismate synthase, a chorismate mutase, a dehydrogenase, e.g., a prephenate dehydrogenase, and an amino transferase. For example, using the *S. Venezuelae* enzymes *PapA*, *PapB*, and *PapC* together with an *E. coli* aminotransferase, e.g., as shown in Figure 15, Panel B, *PapA*, chorismate is used to produce pAF.

[369] For example, 4-amino-4-deoxychorismate synthase converts chorismate to 4-amino-4-deoxychorismic acid (compound 3 in Figure 15, Panel B), e.g., using ammonia (from glutamine) in a simple addition-elimination reaction. *PapB* and *PapC*, which are analogous to chorismate mutase and prephenate dehydrogenase, respectively, are used to convert 4-amino-4-deoxychorismic acid to 4-amino-4-deoxyprephenic acid (compound 4 in Figure 15, Panel B) and then to *p*-aminophenyl-pyruvic acid (compound 5 in Figure 15, panel B). A non-specific tyrosine aminotransferase, e.g., from *E. coli* is used to convert *p*-aminophenyl-pyruvic acid to pAF. See, e.g., *Escherichia coli and Salmonella*, 2nd ed, ed. F.C. Neidhardt. Vol. 1. 1996, Washington, D. C.: ASM Press. For example, *tyrB*, *aspS*, or *ilvE* is optionally used to produce pAF from *p*-aminophenyl-pyruvic acid.

[370] Figure 13 illustrates a plasmid for use in the biosynthesis of pAF. The plasmid depicted comprises *S. Venezuele* genes *papA*, *papB*, and *papC* cloned into a pSC101 derived pLASC plasmid, e.g., under control of the *lac* or *lpp* promoter. The plasmid is used to transform a cell, e.g., a bacterial cell, such that cell produces the enzymes encoded by the genes. When expressed, the enzymes catalyze one or more reactions designed to produce a desired unnatural amino acid, e.g., pAF. For example, proteins *PapA*, *PapB* and *PapC* convert chorismate to *p*-aminophenyl-pyruvic acid, while an *E. coli* aromatic aminotransferase completes the biosynthesis to afford pAF.

[371] Typically, the synthesis of pAF from chorismate, in the present invention does not affect the concentration of other amino acids produced in the cell, e.g., other aromatic amino acids typically produced from chorismate. Typically, *p*-aminophenylalanine is

WO 02/085923

PCT/US02/12465

produced in a concentration sufficient for efficient protein biosynthesis, e.g., a natural cellular amount, but not to such a degree as to affect the concentration of the other aromatic amino acids or exhaust cellular resources. Typical concentrations of pAF produced *in vivo* in this manner are about 10 mM to about 0.05 mM. In *S. Venezuelae* evidence suggests that the regulation of the shikimate pathway is modified to account for chorismate consumption in making a fourth aromatic amino acid. See, e.g., He, J., et al., *Microbiology*, 2001 147: p. 2817-2829. Once a bacterium is transformed with the plasmid comprising the genes used to produce enzymes used in the above pathway, and pAF as a twenty-first amino acid is generated, *in vivo* selections are optionally used to further optimize the production of pAF for both ribosomal protein synthesis and cell growth. [372] Since a pAF tRNA-synthetase pair allows the suppression of a TAG codon in a nonessential position of a protein, biosynthetic pathway effectiveness is optionally monitored and optimized by the production of that protein. Only cells that produce a concentration of pAF sufficient for protein biosynthesis are able to suppress the TAG codon. At the same time, one can select for optimal pAF production based on *E. coli* growth rates if the TAG-protein is an essential protein to cell growth. Placing the biosynthetic genes on a plasmid allows the level of pAF produced to be modified, e.g., by changing plasmid copy number and promotor strength. To determine if the addition of a pAF biosynthetic pathway affects the production of other aromatic amino acids in *E. coli*, and to quantitate pAF production, the cellular concentrations of the aromatic amino acids is optionally monitored, e.g., by extraction and LCMS analysis. See, e.g., Moss, R.E., *Methods in Enzymology*, 1995, 262: p. 497-499 and Mimura, H., S. Nagata, and T. Matsumoto, *Biosci. Biotech. Biochem.*, 1994, 58(10): p. 1873-1874

Example 13: Biosynthesis of Dopa

[373] To biosynthetically produce dopa *in vivo*, one or more genes, e.g., *hpaBC*, for a nonspecific aromatic hydroxylase, e.g., from *E. coli* are cloned into a low copy number vector, e.g., a pSC101 derivative, which is typically placed under control of an *lpp* promoter. This construct produces dopa (2) from tyrosine (1), *in vivo*, as shown in Figure 20 while not being toxic to the growing cells. Similar work was done with this gene to overproduce dopa for purification purposes. See, e.g., Jang-Young Lee, Luying Xun, *Biotechnology Letters*, 1998, 20, 479-482. However, as described above, overproduction is not typically desired. In this application, a low copy plasmid is used to produce dopa in a natural cellular amount.

WO 02/085923

PCT/US02/12465

Example 14: Biosynthesis of O-methyl-L-tyrosine

[374] O-methyl-L-tyrosine is optionally produced biosynthetically by plant O-methyltransferases are enzymes involved in secondary metabolism, which converts a hydroxyl group into a methoxyl group. Two such enzymes were selected: (iso)eugenol O-methyltransferase (IEMT) and caffeic acid O-methyltransferase (COMT). Both of them are from *Clarkia breweri*. IEMT methylates eugenol/isoeugenol, and COMT methylates caffeic acid. The substrates of these two enzymes are similar to tyrosine. However, both enzymes have high substrate specificity and methylation regiospecificity. Therefore, a combinatorial approach to evolve these two enzymes was adopted so that they would take tyrosine as their substrate and convert tyrosine into O-methyl-L-tyrosine. Active site residues were selected for mutation, and large mutant libraries were created. After several rounds of selection, at least about three hits have been identified.

[375] In other embodiments, the enzymes used to produce O-methyl-L-tyrosine can also be artificially evolved, e.g., to produce a meta substituted methoxy phenylalanine as provided in Formula III.

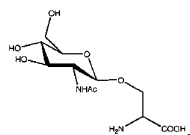
Example 15: Biosynthesis of glycosylated amino acids

[376] The present invention also provides biosynthetic methods for the production of glycosylated amino acids. Forming glycosylated amino acids in vivo is optionally performed in a number of ways. For example, transforming a cell with a plasmid comprising a gene for a N-acetyl-galactosaminidase, a transglycosylase, or a hydrolase, e.g., serine-glycosyl hydrolase, e.g., acting in the reverse direction, provides a cell that produces a glycosylated amino acid. When combined with a translation system as provided below, the biosynthetic pathway results in a cell that produces and incorporates a glycosylated amino acid into one or more proteins within the cell. For example, see, e.g., Figure 28, illustrating the formation of a glycosylated amino acid, wherein R is optionally an alcohol, an amine, or an N-acetyl amine. An example structure is shown by Formula IV:

WO 02/085923

PCT/US02/12465

IV



[377]

Example 16: Identification of advantages due to incorporation of unnatural amino acids

[378] Given the capability presented herein of developing a completely autonomous
 5 bacterium that can biosynthesize a unnatural amino acid from basic carbon sources and
 incorporate this amino acid into proteins, e.g., in response to a nonsense codon in DNA,
 with high translational efficiency and fidelity, the question remains whether such additions
 actually provide an advantage to the bacterium over an organism that incorporates only the
 twenty natural amino acids. The present invention provides a method of determining if an
 10 expanded genetic code provides any such advantage as well as identifying the type of
 advantage and the unnatural amino acid to which it is due.

[379] Since the 19th century bacteriologists have been interested in the extraordinary
 changes of bacterial cultures grown under various conditions. See, e.g., Summers, W.C.,
J. Hist. Biol., 1991, 24: P. 171-190. However, all forms of evolution have been studied
 15 with twenty amino acid organisms. The present invention addresses the feasibility of
 expanding the genetic code of *E. coli* with unnatural amino acids and provides methods of
 testing whether the ability to incorporate additional amino acids provides *E. coli* with an
 evolutionary advantage.

[380] To determine whether the addition of novel amino acids to the genetic code can
 20 provide an evolutionary advantage to *E. coli*, the evolution of a twenty-one amino acid
 bacteria is optionally compared to that of a twenty amino acid bacteria. The approach
 combines new sets of translational machinery for incorporation of unnatural amino acids
 into proteins with a mutagenized *E. coli* genomic library placed under selective pressures.
 The genetic, selection approach described above and elsewhere by Schultz and coworkers
 25 has, thus far, produced at least about eleven new aminoacyl synthetases that can
 incorporate novel amino acid into proteins efficiently and with high fidelity in response to
 the TAG codon.

WO 02/085923

PCT/US02/12465

[381] Mutagenizing a plasmid library of the *E. coli* genome scrambles codons and randomly adds TAG nonsense codons throughout the genome. The new TAG codons can be suppressed by the incorporation of new amino acids into the expressed protein. These bacterial systems are placed under selective pressures to select for enhanced *E. coli* growth. See, e.g., Figure 21. The selected genomic fragments from the library that confer an advantage are optionally isolated and screened for enhanced growth ability when incorporating the other unnatural amino acids and tyrosine in response to TAG codons.

[382] A pSC101 low copy vector (approximately 5 copies/cell) is optionally used to construct a large insert (7-14 kb) *E. coli* genomic library by standard methods known to those of skill in the art. For example, a 600 member pSC101 based *E. coli* genomic library provides complete coverage of the *E. coli* genome and is also compatible with the aminoacyl synthetase, and tRNA plasmids described above. Many mutagens have been studied for their ability to incorporate TAG codons into genes. See, e.g., Miller, J.H., A short course in bacterial genetics. 1992, Plainview: Cold Spring Harbor Laboratory Press.

Multiple mutagenesis methods are optionally used since each mutagen is not completely random in its formation of TAG codons. By mutagenizing the same 600 member *E. coli* genomic library with four different mutagens TAG codons are typically placed in as many sites as possible. Four optional methods include, but are not limited to, UV irradiation, a mutator strain (XL1 red), 4-nitro-quinoline-1-oxide (NQO), and ethylmethane sulfonate (EMS) to mutate the genomic library, and combine them to make one large mutated genomic libraries of >1010 members. These mutation methods all rely primarily on forming point mutation but complement each other in the mechanism of mutagenesis resulting over all in a more even distribution of TAG codons. UV irradiation and the mutator strain generate all base substitutions while NQO and EMS principally cause G:C to A:T transitions. Most of the point mutations generated form codons that code for one of the twenty natural amino acids. Since only about 12.5% of the single point mutations can form a TAG codon large highly mutagenized genomic libraries are needed. This method typically generates a least about 106 mutated copies of each gene with many new randomly placed codons. The genomic library is then typically checked for TAG codon incorporation by sequencing a subset of library members before and after mutagenesis.

[383] To determine which genes might be improved by incorporation of one of the new amino acids any of a variety of selective pressures are optionally used for screening that target a range of cellular biology: catalytic functions, protein interactions, carbon sources, multiple response genes, and broad metabolic functions. For example, selection pressure

WO 02/085923

PCT/US02/12465

- based on quinolones is used to target topoisomerase and DNA gyrase. 5-fluorouracil is used to target DNA synthesis; omeprazole is used to target proton pump inhibitors; the use of fatty acids as a sole carbon source and acidic media are used to target a variety of genes related to utilization of carbon and response; and a reductive media is used to target the
- 5 thiol-redox pathway and disulfide containing proteins. See, e.g., Bronson, J.J. and J.F. Barrett, *Curr. Med. Chemistry*, 2001 8: p. 1775-1793; Bearden, D.T. and L.H. Danziger, *Pharmacotherapy*, 2001 21(10): p. 224S-232S; Matthews, D.A., et al., *J. Mol. Biol.*, 1990. 214(4): p. 937-948; Knox, M.R. and J.E. Harris, *Arch. Microbiol.*, 1988. 149(6): p. 557-60; McGowan, C.C., T.L. Cover, and M.J. Blaser, *Gastroenterology*, 1994. 107(5): p. 1573-8; Clark, D.P. and J.E. Cronan, Two carbon compounds and fatty acids as carbon
- 10 sources. *Escheria coli and Salmonella cellular and molecular biology*, ed. F.C. Neidhardt. Vol. 1. 1996, Washington D. C: ASM press; Slonczewski, J.L. and J.W. Foster, pH-regulated genes and survival at extreme pH. *Escheria coli and Salmonella cellular and molecular biology*, ed. F.C. Neidhardt. Vol. 1. 1996, Washington D. C.: ASM press; and
- 15 Ritz, D. and J. Beckwith, *Annu. Rev. Microbiol.*, 2001. 55: p. 21-48.
- [384]** The screening of the mutated genomic library produces a set of mutated genomic fragments that confer a growth advantage under a certain selection pressure. These fragments are compared to determine if they are the same found in screens with no unnatural amino acid present by restriction mapping and sequencing. Fragments that
- 20 produce a growth enhancement from an unnatural amino acid selection this fragment are optionally re-screened by comparing growth rate with each unnatural amino acid and tyrosine suppressing the TAG codon. See, e.g., Figure 21. This re-screening of selected genomic fragments insures that the unnatural amino acid is the factor in conferring a growth advantage. For fragments that show a selective growth advantage with an
- 25 unnatural amino acid being inserted into TAG codons, the gene(s) that confers an advantage is optionally isolated and identified, e.g., by digestion and subcloning. The protein can be studied to identify how the unnatural amino acid is enhancing cellular function. The enhanced protein is optionally purified and compared to a natural protein with both in vitro and in vivo studies. Standard enzyme techniques are optionally used to
- 30 study protein stability, kinetics, and its interaction with other biosynthetic pathway components.

WO 02/085923

PCT/US02/12465

Example 17: Sequences

SEQ ID #	Sequence	Notes	tRNA or RS
1	CCGCGGTTAGTTCAGAGAGGCGAGAACGGCGGACTCTAAATCCGCATGGCGCTGGTTC AAATCCGGCCCGCGGACCA	<i>M. jannaschii</i> mtRNA ^{Tyr} _{CUA}	tRNA
2	CCGAGGTTAG CCAAGCTCGG CCAACGGCGA CGAGCTCTAA ATCCGTTCTC GTAGGAGTTC GAGGGTTTGA ATCCGTTCCC TGGGACCA	HLAD03; an optimized amber suppressor tRNA	tRNA
3	GCGAGGTTAG CCAAGCTCGG CCAACGGCGA CGAGCTCTCT AATCCGTTCT CGTAGGAGTT GAGGGTTTGG AATCCGTTCCC CTGGACCA	HL325A; an optimized AGGA transhift suppressor tRNA	tRNA
4	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGCGAGGAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCAGATAGGTTTGAACCAAGT GGTAAATACATTTAGGCGATTATCTCCAAATAAAAAGATGATGATTTACCAAAAT GCTGGATTGATATAATATATTTGGCTGATTTACACGCTATTTAAACCGAGAAA GGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATACAAAAAAGTTTGAAGCA ATGGGTTTAAAGGCAAAATATGTTATGGAAGTACTTCCAGCTTGATTAAGGATTAT ACACTGAATGCTATAGATTGGCTTTAAAGACTACCTTAAAGAGCAAGAGGAGT ATGGAACTTTAGCAAGAGAGGATCAAAATCCAAAGTTGCTGAGTTATCTATCCA ATAATGCAGGTTAATGCATTCATTATCTTGGCGTTGATGTTGCAAGTTGAGGGATG GACAGAGAAAAATACACATGTTAGCAAGGAGCTTTTACCAAAAAAGGTTGTTGTT ATTCACACCCCTCTTTAACCGGTTTGGATGGAAGGAGGATGAGTTTTCAAAAA GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGATAAAGAAA GCATACGCGCGCTGAGTTGTTGAGAGAAATCCAAATATGAGATAGCTAAATAC TCTCTGAAATCTCTTTAGCATTAAGAGCCGAGAAAAATTTGGGAGATTGACAG GTTAGTAGCTATGAGGAGTTAGAGAGTTTATTTAAAAATAGGAATTCATCCAAATG GATTTAAAAAATGCTGAGCTGAGAACTTTATAAGATTATAGAGCCAAATTAAGAAAG AGTTATTA	mutant TyrRS (LW116)	RS
5	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGCGAGGAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTGAGATAGGTTTGAACCAAGT GGTAAATACATTTAGGCGATTATCTCCAAATAAAAAGATGATGATTTACCAAAAT GCTGGATTGATATAATATATTTGGCTGATTTACACGCTATTTAAACCGAGAAA GGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATACAAAAAAGTTTGAAGCA ATGGGTTTAAAGGCAAAATATGTTATGGAAGTACTTCCAGCTTGATTAAGGATTAT ACACTGAATGCTATAGATTGGCTTTAAAGACTACCTTAAAGAGCAAGAGGAGT ATGGAACTTTAGCAAGAGAGGATCAAAATCCAAAGTTGCTGAGTTATCTATCCA ATAATGCAGGTTAATGGTTATCTTCTGGCTGATGTTGCAAGTTGAGGGATG ACACTGAATGCTATAGATTGGCTTTAAAGACTACCTTAAAGAGCAAGAGGAGT TTTCAACCCCTCTTTAACCGGTTTGGATGGAAGGAGGATGAGTTTTCAAAAA GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGATAAAGAAAG CATATCTCCAGCTGAGTTGTTTGAAGGAATCCAAATATGAGGATAGTATTAATCT TCTCTGAAATATCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTGACAG TTATATAGCTATGAGGAGTTAGAGAGTTTATTTAAAAATAGGAATTCATCCAAATG ATTTAAAAAATGCTGAGCTGAGAACTTTATAAGATTATAGAGCCAAATTAAGAAAG AGTTA	p-Pr-PheRS	RS
6	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGCGAGGAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCAGATAGGTTTGAACCAAGT GGTAAATACATTTAGGCGATTATCTCCAAATAAAAAGATGATGATTTACCAAAAT GCTGGATTGATATAATATATTTGGCTGATTTACACGCTATTTAAACCGAGAAA GGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATACAAAAAAGTTTGAAGCA ATGGGTTTAAAGGCAAAATATGTTATGGAAGTACTTCCAGCTTGATTAAGGATTAT ACACTGAATGCTATAGATTGGCTTTAAAGACTACCTTAAAGAGCAAGAGGAGT ATGGAACTTTAGCAAGAGAGGATCAAAATCCAAAGTTGCTGAGTTATCTATCCA ATAATGCAGGTTAATGGTTATCTTCTGGCTGATGTTGCAAGTTGAGGGATG ACACTGAATGCTATAGATTGGCTTTAAAGACTACCTTAAAGAGCAAGAGGAGT TTTCAACCCCTCTTTAACCGGTTTGGATGGAAGGAGGATGAGTTTTCAAAAA GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGATAAAGAAAG CATATCTCCAGCTGAGTTGTTTGAAGGAATCCAAATATGAGGATAGTATTAATCT TCTCTGAAATATCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTGACAG TTATATAGCTATGAGGAGTTAGAGAGTTTATTTAAAAATAGGAATTCATCCAAATG ATTTAAAAAATGCTGAGCTGAGAACTTTATAAGATTATAGAGCCAAATTAAGAAAG AGTTA	p-NH ₂ -PheRS(1)	RS
7	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGCGAGGAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCAGATAGGTTTGAACCAAGT GGTAAATACATTTAGGCGATTATCTCCAAATAAAAAGATGATGATTTACCAAAAT GCTGGATTGATATAATATATTTGGCTGATTTACACGCTATTTAAACCGAGAAA GGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATACAAAAAAGTTTGAAGCA ATGGGTTTAAAGGCAAAATATGTTATGGAAGTACTTCCAGCTTGATTAAGGATTAT ACACTGAATGCTATAGATTGGCTTTAAAGACTACCTTAAAGAGCAAGAGGAGT ATGGAACTTTAGCAAGAGAGGATCAAAATCCAAAGTTGCTGAGTTATCTATCCA ATAATGCAGGTTAATGGTTATCTTCTGGCTGATGTTGCAAGTTGAGGGATG ACACTGAATGCTATAGATTGGCTTTAAAGACTACCTTAAAGAGCAAGAGGAGT TTTCAACCCCTCTTTAACCGGTTTGGATGGAAGGAGGATGAGTTTTCAAAAA GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGATAAAGAAAG CATATCTCCAGCTGAGTTGTTTGAAGGAATCCAAATATGAGGATAGTATTAATCT TCTCTGAAATATCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTGACAG TTATATAGCTATGAGGAGTTAGAGAGTTTATTTAAAAATAGGAATTCATCCAAATG ATTTAAAAAATGCTGAGCTGAGAACTTTATAAGATTATAGAGCCAAATTAAGAAAG AGTTA	p-NH ₂ -PheRS(2)	RS

SEQ ID #	Sequence	Notes	tRNA or RS
8	TAAATAGCATAGGAGATAGAGAGATTATTTAAAAAATGAAGGATTCGCATCCATAGGATATTTAAAAAATCGCTAGTGTGAAGAACTTATTAAGAATTATTGAGGCCAATTAAGAAAGA GATTA ATGCGAGCAATTGGAATATTAAGAAGAAACACATCTGAATATTATCAGCGAGGAAGAGT TTAAGAGAGGATTTAAAAAAGAGCAAAATCTGCTCAATATAGGTTTGTACACAAAT GCTGGAATTTGATATACATTTTATTTGTGGCGATTATACACGGATTTTAAACCCAGAAA GGAGAGTGGATAGAGTATAGAAAATTTATGGATATTAACAAATAAAGGTTTGTGAAGCA ATGGGGTTAAAGCGAAATATATGTGTTATGGAGATGATTTCTGATATCTGATATGAAGATAT ACATCGAATCTTATGATTTGGGTTTAAAAATCTGTTAAAGAAAGCAGCAAGAGAGTT ATGGAATCTTAAGAAGAGAGATGAATAATCAAGAGTGTGCTGAGTTGTATCATCTCAA TATCGGATTTATTCGGCGGATATCTGCTGGCGGTGAGTGTGCAGTTGGAGGAGTGG GCGAGAAATATACATCTTTACAGAGAGAGTTTTCACAAATGATTTGTTGATTTGATTA TTCACAACTCTGCTTAAGGGGTTTGGATGGAGAGAGTTTATCCAAAGATTTCTTCAAAG GGATTTTATACGCTGTGATGAGTCTCCAGAGAGAGTTTAGGGCTGTATGATTAAGCAAG ATGCTGCTCCGCGAGAGATCTTTTGAAGGAATCCATATTAAGAGATTTATCTCAAGATAT TCTTGGAATCTTATCAATCAATAAAAGCGCAAAATATTTGCTGGAGTTTGTGACG TATTATAGCTATGACGAGTATGAGAGGTTTATTTAAAAAATGAAGATTCATCTCCATGTG ACTTAAAAAATCGCTTAGCTGAGAGACTTATAAAGATTATTGAGGCCAATTAAGAAAGA GATTA	p-NH ₂ -PheRS(3a)	RS
9	ATGCGAGCAATTGGAATATTAAGAAGAAACACATCTGAATATTATCAGCGAGGAAGAGT TTAAGAGAGGATTTAAAAAAGAGCAAAATCTGCTCAATATAGGTTTGTACACAAAT GCTGGAATTTGATATACATTTATTTGTGGCGATTATACACGGATTTTAAACCCAGAAA GGAGAGTGGATAGAGTATAGAAAATTAAGGAGATTAACAAAATAATTTTGTGAACA ATGGGGTTAAAGCGAAATATGTGTTATGGAGATCTCATATTAAGAGATTTATCTCAAG ATGCGATTTAAAGCAAAATATTTGTTTATGGAGTCTGCTCCAGTTGATAGGATAT ACATCTGAATCTTATAGATTCGGCTTAAAAAATCTTAAAGAAAGCAGCAAGAGGTT ATGGAATCTTAAGAAGAGAGATGAATAATCAAGAGTGTGCTGAGGTTATCTATCCAA TATCGGATTTATGAGAGCATTTATGATGGCGTTGATGTTGATGTTGATTTGATTA ACAGAGAAATATACATCAATTTACAGCAAGAGGTTTATCCAAAGAGTTTCTTTGTA TTCACAACTCTGCTTAAGGGGTTTGGATGGAGAGAGAGATAGGTTTTCCTTCAAAG GGATTTTATACGCTGTGATGAGTCTCCAGAGAGAGTTTAGGGCTGTATGATTAAGCAAG ATGCTGCTCCGCGAGAGATCTTTTGAAGGAATCCATATTAAGAGATTTATCTCAAGAT TCTTGGAATCTTATCAATCAATAAAAGCGCAAAATATTTGCTGGAGTTTGTGACG TAAAGAGATGAGAGATTTATTTAAATAATGAGATTAAGAGATTTTAGGCCAATTAAGAAAGA GATTA	p-NH ₂ -PheRS(3b)	RS
10	ATGCGAGCAATTGGAATATTAAGAAGAAACACATCTGAATATTATCAGCGAGGAAGAGT TTAAGAGAGGATTTAAAAAAGAGCAAAATCTGCTCAATATAGGTTTGTACACAAAT GCTGGAATTTGATATACATTTATTTGTGGCGATTATACACGGATTTTAAACCCAGAAA GGAGAGTGGATAGAGTATAGAAAATTAAGGAGATTAACAAAATAATTTTGTGAACA ATGGGGTTAAAGCGAAATATGTGTTATGGAGATCTCATATTAAGAGATTTATCTCAAG ATGCGATTTAAAGCAAAATATTTTATGTGAGATTTGCTCCAGCTTATGATAGGATTT ACATCTGAATGCTATAGATTCGGCTTAAAAAATCTCTTAAAGAAAGCAGCAAGAGGTT ATGGAATCTTAAGAAGAGAGATGAATAATCAAGAGTGTGCTGAGGTTATCTATCCAA TATCGGATTTATGAGAGCATTTATGATGGCGTTGATGTTGATGTTGATTTGATTA ACAGAGAAATATACATCAATTTACAGCAAGAGGTTTATCCAAAGAGTTTCTTTGTA TTCACAACTCTGCTTAAGGGGTTTGGATGGAGAGAGAGATAGGTTTTCCTTCAAAG GGATTTTATACGCTGTGATGAGTCTCCAGAGAGAGTTTAGGGCTGTATGATTAAGCAAG CATATCTGCCAGCGAGGATGTTTGAAGGAATCCAAATATGAGATATGATTAATCT TCTTGGAATCTTATTTAAACCAATAAAAGCGCAAAATATTTGCTGGAGTTTGTGACG TAAAGAGATGAGAGATTTATTTAAATAATGAGATTAAGAGATTTTAGGCCAATTAAGAAAGA GATTA	O-Allyl-TyrRS(1)	RS
11	ATGCGAGCAATTGGAATATTAAGAAGAAACACATCTGAATATTATCAGCGAGGAAGAGT TTAAGAGAGGATTTAAAAAAGAGTGAUAATCTGCTCATATAGTTTGTGAACAAAT GCTGGAATTTGATATACATTTATTTGTGGCGATTATACACGGATTTTAAACCCAGAAA GGAGAGTGGATAGAGTATAGAAAATTAAGGAGATTAACAAAATAATTTTGTGAAGA ATGGGGTTAAAGCGAAATATGTTTATGGAGATTTGCTCCAGCTTATGATAGGATTT ACATCTGAATGCTATAGATTCGGCTTAAAAAATCTCTTAAAGAAAGCAGCAAGAGGTT ATGGAATCTTAAGAAGAGAGATGAATAATCAAGAGTGTGCTGAGGTTATCTATCCAA TATCGGATTTATGAGAGCATTTATGATGGCGTTGATGTTGATGTTGATTTGATTA ACAGAGAAATATACATCAATTTACAGCAAGAGGTTTATCCAAAGAGTTTCTTTGTA TTCACAACTCTGCTTAAGGGGTTTGGATGGAGAGAGAGATAGGTTTTCCTTCAAAG GGATTTTATACGCTGTGATGAGTCTCCAGAGAGAGTTTAGGGCTTAGATTAAGAAAG CATATCTGCCAGCGAGGATGTTTGAAGGAATCCAAATATGAGAGATATGATTAATCT TCTTGGAATCTTATTTAAACCAATAAAAGCGCAAAATATTTGCTGGAGTTTGTGACG TAAAGAGATGAGAGATTTATTTAAATAATGAGATTAAGAGATTTTAGGCCAATTAAGAAAGA GATTA	O-Allyl-TyrRS(3)	RS
12	ATGCGAGCAATTGGAATATTAAGAAGAAACACATCTGAATATTATCAGCGAGGAAGAGT TTAAGAGAGGATTTAAAAAAGAGATGAUAATCTGCTCATATAGTTTGTGAACAAAT GCTGGAATTTGATATACATTTATTTGTGGCGATTATACACGGATTTTAAACCCAGAAA GGAGAGTGGATAGAGTATAGAAAATTAAGGAGATTAACAAAATAATTTTGTGAAGA ATGGGGTTAAAGCGAAATATGTTTATGGAGATTTGCTCCAGCTTATGATAGGATTT ACATCTGAATGCTATAGATTCGGCTTAAAAAATCTCTTAAAGAAAGCAGCAAGAGGTT ATGGAATCTTAAGAAGAGAGATGAATAATCAAGAGTGTGCTGAGGTTATCTATCCAA TATCGGATTTATGAGAGCATTTATGATGGCGTTGATGTTGATGTTGATTTGATTA ACAGAGAAATATACATCAATTTACAGCAAGAGGTTTATCCAAAGAGTTTCTTTGTA TTCACAACTCTGCTTAAGGGGTTTGGATGGAGAGAGAGATAGGTTTTCCTTCAAAG GGATTTTATACGCTGTGATGAGTCTCCAGAGAGAGTTTAGGGCTTAGATTAAGAAAG CATATCTGCCAGCGAGGATGTTTGAAGGAATCCAAATATGAGAGATATGATTAATCT TCTTGGAATCTTATTTAAACCAATAAAAGCGCAAAATATTTGCTGGAGTTTGTGACG TAAAGAGATGAGAGATTTATTTAAATAATGAGATTAAGAGATTTTAGGCCAATTAAGAAAGA GATTA	O-Allyl-TyrRS(4)	RS

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
	TTCAACACCTGCTTTACCGGTTTGGATGGAGAGGAAAGATGAGTTCTTCAAAAG GGAATTTTATAGCTGTGATGACTCTCCAGAGAGATTAAGGCTATAGATAAAGAAAG CATCTGCCACGCTGAGTTTGTGAAGAAATCCATTAAGAGATAGTAAATATCT TCCTTGAATATCCCTTTACCATATAAAGGCCAGAAAAATTTCTGGAGATTTGACAG TTAATAGCTATGAGAGTTAGAGAGTTTATTTAAAAATAGGATTCATCCAAATGG ATTTAAAAATGCTGTAGCTGAAGACTTATAAAGATTTTAGAGCCAAATTAGAAAGA GATTA		
13	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTTATCAGCGAGGAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCAATATAGTTTGAACCAAGT GCTAAATACATTTTGGGCTTATCTCCAAATAAAGAGATCTGGATTTACAAAAT GCTGGATTTGATATAATATATTTGTTGGCTGATTTACACGCTATTTAAACAGAAA GGAGAGTTGGATGAGATTTAGAAAAATAGGAGATTATACAAAJAAGTTTGTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAGTAAAGTCCAGTTTGTATAGGATTA ACACTGAATGCTATATAGATTGGCTTTAAAAATACCTTAAAGAGCAAGAGGAGT ATGGAACCTTATAGAGAGAGGATGAATCCAAAGGTTGGTGAAGTTATCTATCCAA TAATGCAAGTTATCTGGCTCATTAAGAGGCTTTGATGGAGTTGGAGGCAATGG AGCAGAGAAAAATACACATGTTAGCAAGGGAGCTTTACCAAAAGGTTGTTTGT TTCAACACCTGCTTTACCGGTTTGGATGGAGAGGAAAGATAGTTCTTCAAAAG GGAATTTTATAGCTGTGATGACTCTCCAGAGAGATTAAGGCTATAGATAAAGAGAG CATCTGCCACGCTGGAGTTTGTGAAGAAATCCAAATTAAGGAGATGCTAAATAC TCCTTGAATATCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTGACAG TTAATAGCTATGAGAGATTAGAGAGTTTATTTAAAAATAGGAAATTCATCCAAATGG ATTTAAAAATGCTGTAGCTGAAGACTTATAAAGATTTTAGAGCCAAATTAGAAAGA GATTA	p-Br-PheRS	RS
14	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTTATCAGCGAGGAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCAATATAGTTTGAACCAAGT GCTAAATACATTTTGGGCTATATCTCCAAATAAAGAGATCTGGATTTACAAAAT GCTGGATTTGATATAATATATTTGTTGGCTGATTTACACGCTATTTAAACAGAAA GGAGAGTTGGATGAGATTTAGAAAAATAGGAGATTATACAAAJAAGTTTGTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAGTAAAGTCCAGTTTGTATAGGATTA ACACTGAATGCTATATAGATTGGCTTTAAAAATACCTTAAAGAGCAAGAGGAGT ATGGAACCTTATAGAGAGAGGATGAATCCAAAGGTTGGTGAAGTTATCTATCCAA TAATGCAAGTTATCTGGCTCATTAAGAGGCTTTGATGGAGTTGGAGGCAATGG AGCAGAGAAAAATACACATGTTAGCAAGGGAGCTTTACCAAAAGGTTGTTTGT TTCAACACCTGCTTTACCGGTTTGGATGGAGAGGAAAGATAGTTCTTCAAAAG GGAATTTTATAGCTGTGATGACTCTCCAGAGAGATTAAGGCTTAAAGATAAGAGAG CATCTGCCACGCTGGAGTTTGTGAAGAAATCCAAATTAAGGAGATGCTAAATAC TCCTTGAATATCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTGACAG TTAATAGCTATGAGAGATTAGAGAGTTTATTTAAAAATAGGAAATTCATCCAAATGG ATTTAAAAATGCTGTAGCTGAAGACTTATAAAGATTTTAGAGCCAAATTAGAAAGA GATTA	p-Az-PheRS(1)	RS
15	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTTATCAGCGAGGAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCAATATAGTTTGAACCAAGT GCTAAATACATTTTGGGCTATATCTCCAAATAAAGAGATGATGATTTACAAAAT GCTGGATTTGATATAATATATTTGTTGGCTGATTTACACGCTATTTAAACAGAAA GGAGAGTTGGATGAGATTTAGAAAAATAGGAGATTATACAAAJAAGTTTGTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAGTAAAGTCCAGTTTGTATAGGATTA ACACTGAATGCTATATAGATTGGCTTTAAAAATACCTTAAAGAGCAAGAGGAGT ATGGAACCTTATAGAGAGAGGATGAATCCAAAGGTTGGTGAAGTTATCTATCCAA TAATGCAAGTTATCTGGCTCATTAAGAGGCTTTGATGGAGTTGGAGGCAATGG AGCAGAGAAAAATACACATGTTAGCAAGGGAGCTTTACCAAAAGGTTGTTTGT TTCAACACCTGCTTTACCGGTTTGGATGGAGAGGAAAGATAGTTCTTCAAAAG GGAATTTTATAGCTGTGATGACTCTCCAGAGAGATTAAGGCTTAAAGATAAGAGAG CATCTGCCACGCTGGAGTTTGTGAAGAAATCCAAATTAAGGAGATGCTAAATAC TCCTTGAATATCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTGACAG TTAATAGCTATGAGAGATTAGAGAGTTTATTTAAAAATAGGAAATTCATCCAAATGG ATTTAAAAATGCTGTAGCTGAAGACTTATAAAGATTTTAGAGCCAAATTAGAAAGA GATTA	p-Az-PheRS(3)	RS
16	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTTATCAGCGAGGAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCAATATAGTTTGAACCAAGT GCTAAATACATTTTGGGCTATATCTCCAAATAAAGAGATGATGATTTACAAAAT GCTGGATTTGATATAATATATTTGTTGGCTGATTTACACGCTATTTAAACAGAAA GGAGAGTTGGATGAGATTTAGAAAAATAGGAGATTATACAAAJAAGTTTGTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAGTAAAGTCCAGTTTGTATAGGATTA ACACTGAATGCTATATAGATTGGCTTTAAAAATACCTTAAAGAGCAAGAGGAGT ATGGAACCTTATAGAGAGAGGATGAATCCAAAGGTTGGTGAAGTTATCTATCCAA TAATGCAAGTTATCTGGCTCATTAAGAGGCTTTGATGGAGTTGGAGGCAATGG AGCAGAGAAAAATACACATGTTAGCAAGGGAGCTTTACCAAAAGGTTGTTTGT TTCAACACCTGCTTTACCGGTTTGGATGGAGAGGAAAGATAGTTCTTCAAAAG GGAATTTTATAGCTGTGATGACTCTCCAGAGAGATTAAGGCTTAAAGATAAGAGAG CATCTGCCACGCTGGAGTTTGTGAAGAAATCCAAATTAAGGAGATGCTAAATAC TCCTTGAATATCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTGACAG TTAATAGCTATGAGAGATTAGAGAGTTTATTTAAAAATAGGAAATTCATCCAAATGG ATTTAAAAATGCTGTAGCTGAAGACTTATAAAGATTTTAGAGCCAAATTAGAAAGA GATTA	p-Az-PheRS(5)	RS
17	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTTATCAGCGAGGAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCAATATAGTTTGAACCAAGT GCTAAATACATTTTGGGCTATATCTCCAAATAAAGAGATGATGATTTACAAAAT GCTGGATTTGATATAATATATTTGTTGGCTGATTTACACGCTATTTAAACAGAAA GGAGAGTTGGATGAGATTTAGAAAAATAGGAGATTATACAAAJAAGTTTGTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAGTAAAGTCCAGTTTGTATAGGATTA ACACTGAATGCTATATAGATTGGCTTTAAAAATACCTTAAAGAGCAAGAGGAGT ATGGAACCTTATAGAGAGAGGATGAATCCAAAGGTTGGTGAAGTTATCTATCCAA TAATGCAAGTTATCTGGCTCATTAAGAGGCTTTGATGGAGTTGGAGGCAATGG AGCAGAGAAAAATACACATGTTAGCAAGGGAGCTTTACCAAAAGGTTGTTTGT TTCAACACCTGCTTTACCGGTTTGGATGGAGAGGAAAGATAGTTCTTCAAAAG GGAATTTTATAGCTGTGATGACTCTCCAGAGAGATTAAGGCTTAAAGATAAGAGAG CATCTGCCACGCTGGAGTTTGTGAAGAAATCCAAATTAAGGAGATGCTAAATAC TCCTTGAATATCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTGACAG TTAATAGCTATGAGAGATTAGAGAGTTTATTTAAAAATAGGAAATTCATCCAAATGG ATTTAAAAATGCTGTAGCTGAAGACTTATAAAGATTTTAGAGCCAAATTAGAAAGA GATTA	Mutant synthetases to incorporate m-acyl phenylalanine into proteins (Ketone 3-4)	RS

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	IRNA or RS
	ACACTGAATGCTATAGATTGGCTTTAAAAAATACCTTAAAAAGAGCAAGAGAGT ATGGAACCTTATAGCAAGAGAGGATGAAAAATCCAAAGGTTGCTGAAGTTATCTATCC ATAATGACGGTTATAGGAATCAATATCAAGGCTTGATATGCTGGAGGAGT GAGCAGAGAAAAATACACATCTTAGCAAGGGAGCTTTACCAAAAAAGGTTGTTGT ATTCACAAACCTGCTTTAACGGGTTTGGATGGAGAGGAAAGATGAGTTCTTCAAAA GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGTAAGAAA GCTACCTCCAGCTGGATTTTGAGAGAAATCCAAATATGAGATAGCTAAATAC TTCCCTGAATATCCTTTAACCATAAAAAGGCTAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGCTTTATTTAAAAATAGGAATGCTATCCAAATG GATTTAAAAATGCTTTAGCTGAGGACTTATAAAGAATTTAGAGCAATTTAGAAAG AGATTATAA		
18	ATGAGCGAATTTGAAATGATAAAGAGAAACACATCTGMAATTTATCAGCAGGAAAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCTAATAGGTTTGAACCAAGT GGTAATAATACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTATACAAAAAT OCTGGATTGATATAATATATTTGCTGGCTGATTTACAGGCTATTAAACAGAGAAA GGAGGTTGGATGAGTTAGAAAAATGAGGATTAACAAAAAAGTTTTTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAAGCTTATTCAGCTTGATAAGGATTAT ACACTGAATGCTATAGATTGGCTTTAAAAAATACCTTAAAAAGAGCAAGAGAGT ATGGAACCTTATAGCAAGAGAGGATGAAAAATCCAAAGGCTGCTGAGATTATCTTACA ATAATGACGGTTATAGGAATCAATATACAGGCTTGATTTGAGCTTGAAGGAGT GAGCAGAGAAAAATACACATCTTAGCAAGGGAGCTTTACCAAAAAAGGTTGTTGT ATTCACAAACCTGCTTTAACGGGTTTGGATGGAGAGGAAAGATGAGTTCTTCAAAA GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGTAAGAAA GCTACCTCCAGCTGGATTTTGAAGGAAATCCAAATATGAGATAGCTAAATAC TTCCCTGAATATCCTTTAACCATAAAAAGGCTAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGCTTTATTTAAAAATAGGAATGCTATCCAAATG GATTTAAAAATGCTTTAGCTGAGGACTTATAAAGAATTTAGAGCAATTTAGAAAG AGATTATAA	Mutant synthetase to incorporate m-acyl phenylalanine into proteins (Ketone 3-7)	RS
19	ATGAGCGAATTTGAAATGATAAAGAGAAACACATCTGMAATTTATCAGCAGGAAAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCTAATAGGTTTGAACCAAGT GGTAATAATACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTATACAAAAAT OCTGGATTGATATAATATATTTGCTGGCTGATTTACAGGCTATTAAACAGAGAAA GGAGGTTGGATGAGTTAGAAAAATGAGGATTAACAAAAAAGTTTTTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAAGCTTATTCAGCTTGATAAGGATTAT ACACTGAATGCTATAGATTGGCTTTAAAAAATACCTTAAAAAGAGCAAGAGAGT ATGGAACCTTATAGCAAGAGAGGATGAAAAATCCAAAGGCTGCTGAGATTATCTTACA ATAATGACGGTTATAGGAATCAATATACAGGCTTGATTTGAGCTTGAAGGAGT GAGCAGAGAAAAATACACATCTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTGT ATTCACAAACCTGCTTTAACGGGTTTGGATGGAGAGTTCAGCTTGATAAGGAGT GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGTAAGAAA GCTACCTCCAGCTGGATTTTGAAGGAAATCCAAATATGAGATAGCTAAATAC TTCCCTGAATATCCTTTAACCATAAAAAGGCTAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGCTTTATTTAAAAATAGGAATGCTATCCAAATG GATTTAAAAATGCTTTAGCTGAGGACTTATAAAGAATTTAGAGCAATTTAGAAAG AGATTATAA	Mutant synthetase to incorporate m-acyl phenylalanine into proteins (Ketone 4-1)	RS
20	ATGAGCGAATTTGAAATGATAAAGAGAAACACATCTGMAATTTATCAGCAGGAAAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCTAATAGGTTTGAACCAAGT GGTAATAATACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTATACAAAAAT OCTGGATTGATATAATATATTTGCTGGCTGATTTACAGGCTATTAAACAGAGAAA GGAGGTTGGATGAGATTAGAAAAATAGGAGATTATACAAAAAAGTTTTTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAAGCTTATTCAGCTTGATAAGGATTAT ACACTGAATGCTATAGATTGGCTTTAAAAAATACCTTAAAAAGAGCAAGAGAGT ATGGAACCTTATAGCAAGAGAGGATGAAAAATCCAAAGGCTGCTGAGATTATCTTACA ATAATGCTAGTTAATGATTAATCAATTTTACGGCTTGATTTGATTTGAGGAGGATG GAGCAGAGAAAAATACACATCTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTGT ATTCACAAACCTGCTTTAACGGGTTTGGATGGAGAGTTCAGCTTGATAAGGAGT GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGTAAGAAA GCTACCTCCAGCTGGATTTTGAAGGAAATCCAAATATGAGATAGCTAAATAC TTCCCTGAATATCCTTTAACCATAAAAAGGCTAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGCTTTATTTAAAAATAGGAATGCTATCCAAATG GATTTAAAAATGCTTTAGCTGAGGACTTATAAAGAATTTAGAGCAATTTAGAAAG AGATTATAA	Mutant synthetase to incorporate m-acyl phenylalanine into proteins (Ketone 5-4)	RS
21	ATGAGCGAATTTGAAATGATAAAGAGAAACACATCTGMAATTTATCAGCAGGAAAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCTAATAGGTTTGAACCAAGT GGTAATAATACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTATACAAAAAT OCTGGATTGATATAATATATTTGCTGGCTGATTTACAGGCTATTAAACAGAGAAA GGAGGTTGGATGAGATTAGAAAAATAGGAGATTATACAAAAAAGTTTTTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAAGCTTATTCAGCTTGATAAGGATTAT ACACTGAATGCTATAGATTGGCTTTAAAAAATACCTTAAAAAGAGCAAGAGAGT ATGGAACCTTATAGCAAGAGAGGATGAAAAATCCAAAGGCTGCTGAGATTATCTTACA ATAATGCTAGTTAATGATTAATCAATTTTACGGCTTGATTTGATTTGAGGAGGATG GAGCAGAGAAAAATACACATCTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTGT ATTCACAAACCTGCTTTAACGGGTTTGGATGGAGAGGAAAGATGAGTTCTTCAAAA GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGTAAGAAA GCTACCTCCAGCTGGATTTTGAAGGAAATCCAAATATGAGATAGCTAAATAC TTCCCTGAATATCCTTTAACCATAAAAAGGCTAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGCTTTATTTAAAAATAGGAATGCTATCCAAATG GATTTAAAAATGCTTTAGCTGAGGACTTATAAAGAATTTAGAGCAATTTAGAAAG AGATTATAA	Mutant synthetase to incorporate m-acyl phenylalanine into proteins (Ketone 6-8)	RS
22	ATGAGCGAATTTGAAATGATAAAGAGAAACACATCTGMAATTTATCAGCAGGAAAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTCTAATAGGTTTGAACCAAGT GGTAATAATACATTTAGGGCATTATCTCCAAATAAAAAAGATGATTATACAAAAAT OCTGGATTGATATAATATATTTGCTGGCTGATTTACAGGCTATTAAACAGAGAAA GGAGGTTGGATGAGATTAGAAAAATAGGAGATTATACAAAAAAGTTTTTGAAGCA ATGGGTTTAAAGCAAAATATGTTTATGGAAGCTTATTCAGCTTGATAAGGATTAT ACACTGAATGCTATAGATTGGCTTTAAAAAATACCTTAAAAAGAGCAAGAGAGT ATGGAACCTTATAGCAAGAGAGGATGAAAAATCCAAAGGCTGCTGAGATTATCTTACA ATAATGCTAGTTAATGATTAATCAATTTTACGGCTTGATTTGATTTGAGGAGGATG GAGCAGAGAAAAATACACATCTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTGT ATTCACAAACCTGCTTTAACGGGTTTGGATGGAGAGGAAAGATGAGTTCTTCAAAA GGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTTAAGTAAGAAA GCTACCTCCAGCTGGATTTTGAAGGAAATCCAAATATGAGATAGCTAAATAC TTCCCTGAATATCCTTTAACCATAAAAAGGCTAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGCTTTATTTAAAAATAGGAATGCTATCCAAATG GATTTAAAAATGCTTTAGCTGAGGACTTATAAAGAATTTAGAGCAATTTAGAAAG AGATTATAA	Mutant synthetase to incorporate m-	RS

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
	GOTAAAAATACATTAGGCGATATCTCCAAATAAAAAAGATGATTGATTACAAAAAT GCTGGATTGATATAATTAATATTGTGTGCTGATTACACGCTATTTAACCCAGAAA GGAGAGTTGGATGCTGTTGAAAAATGAGGATTTACAAAAAGAGTTTGTGAAGCA ATGGGGTTAAAGCCAAAATATGTTTATGGAAGTGAATCCAGCTTGATAAGGATTAT ACACTGAATGCTACAGATTGCTTTAAAAAATACCTTAAAGAGGCAAGAGAGAT ATGGAACCTTATAGCAAGAGGCTGAAAAATGCAAGGTTCTGAGATTATCTTCCCA ATCATCCGCTTAATCATGATTCAATTATGCAAGGCTGATGTTGAGTTGAGGGATG GAGCAGAGAAAAATACACATGTTAGCAAGGAGCTTTTACCAAAAAAGGTTGTTGT ATTACAAACCTGCTTTACCGGCTTTGGATGGAGAGGAAAGATGATTCTTCAAAA GGCAATTTTATGCTGTGGAGCTCTTCCAGAGAGATTTAGGCTTAAGCTTAAGAAA GCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCAAATATGGAGATAGCTAAATAC TTCTTGAAATGCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTACAGAGTTTATTAATAATAGGAAATGCTATCCATC GATTAAAAAATGCTGTAGCTGAGAACCTTATAAGATTTTAGAGCCAAATAGAAAAG AGATTATAA	methoxy phenylalanine into proteins (OMe 1-6)	
23	ATGGAGCAATTGAAATGATAAAGAGAACACATCTGAAATATACAGCGAGGAAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTACAAATAGGTTTGAACCAAGT GOTAAAAATACATTAGGCGATATCTCCAAATAAAAAAGATGATTGATTACAAAAAT GCTGGATTGATATAATTAATATTGTGTGCTGATTACACGCTATTTAACCCAGAAA GGAGAGTTGGATGCTGTTGAAAAATGAGGATTTACAAAAAAGTTTGTGAAGCA ATGGGGTTAAAGCCAAAATATGTTTATGGAAGTGAATCCAGCTTGATAAGGATTAT ACACTGAATGCTACAGATTGCTTTAAAAAATACCTTAAAGAGGCAAGAGAGAT ATGGAACCTTATAGCAAGAGGCTGAAAAATGCAAGGTTCTGAGATTATCTTCCCA ATCATCCGCTTAATCATGATTCAATTATGAGCTGATGTTGAGTTGAGGGATG GAGCAGAGAAAAATACACATGTTAGCAAGGAGCTTTTACCAAAAAAGGTTGTTGT ATTACAAACCTGCTTTACCGGCTTTGGATGGAGAGGAAAGATGATTCTTCAAAA GGCAATTTTATGCTGTGATGACTCTCCAGAGAGATTAAGGCTTAAGATAAGAAA GCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCAAATATGGAGATAGCTAAATAC TTCTTGAAATGCTTTAACCAATAAAAGGCCAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGTTTATTTAAAAATAGGAATTCATCCAAATG GATTAAAAAATGCTGTAGCTGAGAACCTTATAAGATTTTAGAGCCAAATAGAAAAG AGATTATAA	Mutant synthetases to incorporate m- methoxy phenylalanine into proteins (OMe 1-8)	RS
24	ATGGAGCAATTGAAATGATAAAGAGAACACATCTGAAATATACAGCGAGGAAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTACAAATAGGTTTGAACCAAGT GOTAAAAATACATTAGGCGATATCTCCAAATAAAAAAGATGATTGATTACAAAAAT GCTGGATTGATATAATTAATATTGTGTGCTGATTACACGCTATTTAACCCAGAAA GGAGAGTTGGATGCTGATTAGAAAAATAGGAGATTATACAAAAAAGTTTGTGAAGCA ATGGGGTTAAAGCCAAAATATGTTTATGGAAGTGAATCCAGCTTGATAAGGATTAT ACACTGAATGCTACAGATTGCTTTAAAAAATACCTTAAAGAGGCAAGAGAGAT ATGGAACCTTATAGCAAGAGGAGTGAATAATCCAAAGTTGCTGAGTTATCTATCCA ATAATGCAAGTTTAAATCATACATTAAGACGCTGATGCTGAGGTTGGAGGGATG GAGCAGAGAAAAATACACATGTTAGCAAGGAGCTTTTACCAAAAAAGGTTGTTGT ATTACAAACCTGCTTTACCGGCTTTGGATGGAGAGGAAAGATGATTCTTCAAAA GGCAATTTTATGCTGTGATGACTCTCCAGAGAGATTAAGGCTTAAGATAAGAAA GCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCAAATATGGAGATAGCTAAATAC TTCTTGAAATGCTTTAACCAATAAAAGGCCAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGTTTATTTAAAAATAGGAATTCATCCAAATG GATTAAAAAATGCTGTAGCTGAGAACCTTATAAGATTTTAGAGCCAAATAGAAAAG AGATTATAA	Mutant synthetase to incorporate m- methoxy phenylalanine into proteins (OMe 2-7)	RS
25	ATGGAGCAATTGAAATGATAAAGAGAACACATCTGAAATATACAGCGAGGAAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTACAAATAGGTTTGAACCAAGT GOTAAAAATACATTAGGCGATATCTCCAAATAAAAAAGATGATTGATTACAAAAAT GCTGGATTGATATAATTAATATTGTGTGCTGATTACACGCTATTTAACCCAGAAA GGAGAGTTGGATGCTGATTAGAAAAATAGGAGATTATACAAAAAAGTTTGTGAAGCA ATGGGGTTAAAGCCAAAATATGTTTATGGAAGTGAATCCAGCTTGATAAGGATTAT ACACTGAATGCTACAGATTGCTTTAAAAAATACCTTAAAGAGGCAAGAGAGAT ATGGAACCTTATAGCAAGAGGAGTGAATAATCCAAAGTTGCTGAGTTATCTATCCA ATAATGCAAGTTTAAATCATACATTAAGACGCTGATGCTGAGGTTGGAGGGATG GAGCAGAGAAAAATACACATGTTAGCAAGGAGCTTTTACCAAAAAAGGTTGTTGT ATTACAAACCTGCTTTACCGGCTTTGGATGGAGAGGAAAGATGATTCTTCAAAA GGCAATTTTATGCTGTGATGACTCTCCAGAGAGATTAAGGCTTAAGATAAGAAA GCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCAAATATGGAGATAGCTAAATAC TTCTTGAAATGCTTTAACCAATAAAAGGCCAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGTTTATTTAAAAATAGGAATTCATCCAAATG GATTAAAAAATGCTGTAGCTGAGAACCTTATAAGATTTTAGAGCCAAATAGAAAAG AGATTATAA	Mutant synthetase to incorporate m- methoxy phenylalanine into proteins (OMe 4-1)	RS
26	ATGGAGCAATTGAAATGATAAAGAGAACACATCTGAAATATACAGCGAGGAAGAG TTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTACAAATAGGTTTGAACCAAGT GOTAAAAATACATTAGGCGATATCTCCAAATAAAAAAGATGATTGATTACAAAAAT GCTGGATTGATATAATTAATATTGTGTGCTGATTACACGCTATTTAACCCAGAAA GGAGAGTTGGATGCTGATTAGAAAAATAGGAGATTATACAAAAAAGTTTGTGAAGCA ATGGGGTTAAAGCCAAAATATGTTTATGGAAGTGAATCCAGCTTGATAAGGATTAT ACACTGAATGCTACAGATTGCTTTAAAAAATACCTTAAAGAGGCAAGAGAGAT ATGGAACCTTATAGCAAGAGGAGTGAATAATCCAAAGTTGCTGAGTTATCTATCCA ATAATGCAAGTTTAAATCATACATTAAGACGCTGATGCTGAGGTTGGAGGGATG GAGCAGAGAAAAATACACATGTTAGCAAGGAGCTTTTACCAAAAAAGGTTGTTGT ATTACAAACCTGCTTTACCGGCTTTGGATGGAGAGGAAAGATGATTCTTCAAAA GGCAATTTTATGCTGTGATGACTCTCCAGAGAGATTAAGGCTTAAGATAAGAAA GCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCAAATATGGAGATAGCTAAATAC TTCTTGAAATGCTTTAACCAATAAAAGGCCAGAAAAATTTGGTGGAGATTGACA GTTAATAGCTATGAGGAGTTAGAGGTTTATTTAAAAATAGGAATTCATCCAAATG GATTAAAAAATGCTGTAGCTGAGAACCTTATAAGATTTTAGAGCCAAATAGAAAAG AGATTATAA	Mutant synthetase to incorporate m- methoxy phenylalanine into proteins (OMe 4-8)	RS

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
27	GATTAAAAATGCTGTAGCTGAAGAATCTATAAGATTTTAGMGCCAAATAGAAAGAGATTATAA ATGACGAAATTTGAATGATAAAGAGAAACACATCTGAAATTTACAGCGAGGAAGAGTTAAAGAGAGCTTTAAAAAAGATGAAATAATCTGCTTACATAGGTTTGAACCAAGTGGTAAAAATACATTTAGGCGATTATCTCCAAATAAAAAGATGATTGATTACAAAATGCTGGATTGGATATAATATATTTGGCTGATTTACACGCCATTAAACAGAGAAAGGAGTTGGATGAGCTGGAGAAATATGGAATATGAGGTTTATACAAAAGTTTGAAGCAATGGGTTAAAGGCAAAATATGTTTATGGAAGTGCAATCCAGCTTGATAGGATTATACACTGAATGCTATGATTTGGCTTTAAAAACTACCTTAAAAAGAGCAAGAGGAGTATGGAATTTTAGCAAGAGGAGGAGAAATCCAAAGGTTCTCAAGTTATCCATCCATAAATGCAAGTTAAATGCGCACATTTATTTAGGCGTTGATGTTGACGTTGAGGGATGGAGCAGAGAAAAATACACATGTTTAGCAAGGAGCTTTACCAAAAAGGTTGTTTGTATCCACACCCCTGCTTACCGGTTTGGATGGAGAGGAGATGAGGATTTCTCAAAAAGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTAAGATAAAGAAAGCACTAGCCAGCTGGAGTTGTTGAGAGAAATCCAAATATGAGATAGCTAAATACCTCTTGATATCTTACCATTAAGAGGCGAGAAATTTGTGAGATTTGACAATTAAATAGCTATGAGGAGTTAGAGATTTATTTAAAAATAGGAATTCGATCCAAATGATTAAAAAATGCTGTAGCTGAAGAATCTATAAGATTTTAGAGCCAAATAGAAAGAGATTATAA	Mutant synthetase to incorporate p-O-allyl tyrosine into proteins Allyl	RS
28	ATGACGAAATTTGAATGATAAAGAGAAACACATCTGAAATTTACAGCGAGGAAGGTTAAAGAGAGCTTTAAAAAAGATGAAATAATCTGCTTACATAGGTTTGAACCAAGTGGTAAAAATACATTTAGGCGATTATCTCCAAATAAAAAGATGATTGATTACAAAATGCTGGATTGGATATAATATATTTGGCTGATTTACACGCCATTAAACAGAGAAAGGAGTTGGATGAGATTAGAAAAATAGGAGATTATACAAAAGTTTGAAGCAATGGGTTAAAGGCAAAATATGTTTATGGAAGTTGCTTCCAGCTTGATAGGATTATACACTGAATGCTTATAGATTTGGCTTTAAATACCTTAAAGAGGCAAGAGAGGATATGGAATTTTAGAAGAGAGGATGAAATCCAAAGGTTGCTGAAGTTATCTATCCAAATATGCGAGTTTAACTCAGCTTATATCTGAGGTTGATTTCCAGTTGAGGAGATGGAGCAGAGAAATACACATGTTTAGCAGGAGGCTTTTACCAAAAAGGTTGTTTGTATTCACACCCCTGCTTAAACGTTTGGATGGAGAGGAGATGAGCTTCTCAAAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTAAGATAAAGAAAGCATACTGCCAGCTGGAGTTTGAAGAAATCCAAATATGAGATAGCTAAATACCTCTCTGATATCTTTAACCAATAAAGGCCAGAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAGTTAGAGATTTATTTAAAAATAGGAATTCGATCCAAATGATTAAAAAATGCTGTAGCTGAAGAATCTATAAGATTTTAGAGCCAAATAGAAAGAGATTATAA	Aminoacyl tRNA synthetase clone for the incorporation of p-benzoyl-L-phenylalanine (p-BpaRS(H6))	RS
29	ATGGACGAATTTGAATGATAAAGAGAAACACATCTGAAATTTACAGCGAGGAAGGTTAAAGAGAGCTTTAAAAAAGATGAAATAATCTGCTACGATAGGTTTGAACCAAGTGGTAAAAATACATTTAGGCGATTATCTCCAAATAAAAAGATGATTGATTACAAAATGCTGGATTGGATATAATATATTTGGCTGATTTACACGCCATTAAACAGAGAAAGGAGATTGGATGAGATTAGAAAAATAGGAGATTATACAAAAGTTTGAAGCAATGGGTTAAAGGCAAAATATGTTTATGGAAGTTAACTTACAGCTTGATAGGATTATACACTGAATGCTTATAGATTTGGCTTTAAAAATACCTTAAAGAGGCAAGAGAGGATATGGAATTTTAGAAGAGAGGATGAAATCCAAAGGTTGCTGAAGTTATCTATCCAAATATGCGAGTTTAACTCAGCTTATATCTGAGGCTTGAATTTGAGGAGATGGAGCAGAGAAATACACATGTTTACCAAGGAGCTTTTACCAAAAAGGTTGTTTGTATTCACACCCCTGCTTAAACGTTTGGATGGAGAGGAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTAAGATAAAGAAAGCATACTGCCAGCTGGAGTTTGAAGAAATCCAAATATGAGATAGCTAAATACCTCTCTGATATCTTTAACCAATAAAGGCCAGAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAGTTAGAGATTTATTTAAAAATAGGAATTCGATCCAAATGATTAAAAAATGCTGTAGCTGAAGAATCTATAAGATTTTAGAGCCAAATAGAAAGAGATTATAA	Aminoacyl tRNA synthetase clone for the incorporation of p-azido-phenylalanine (p-Az-PheRS(3))	RS
30	ATGGACGAATTTGAATGATAAAGAGAAACACATCTGAAATTTACAGCGAGGAAGGTTAAAGAGAGCTTTAAAAAAGATGAAATAATCTGCTACGATAGGTTTGAACCAAGTGGTAAAAATACATTTAGGCGATTATCTCCAAATAAAAAGATGATTGATTACAAAATGCTGGATTGGATATAATATATTTGGCTGATTTACACGCCATTAAACAGAGAAAGGAGATTGGATGAGATTAGAAAAATAGGAGATTATACAAAAGTTTGAAGCAATGGGTTAAAGGCAAAATATGTTTATGGAAGTTAACTTCCAGCTTGATAGGATTATACACTGAATGCTTATAGATTTGGCTTTAAAAACTACCTTAAAGAGGCAAGAGAGGATATGGAATTTTAGAAGAGAGGATGAAATCCAAAGGTTGCTGAAGTTATCTATCCAAATATGCGAGTTTATCCCTCTTATTAAGAGGCTTGAATTTGCAAGTTGAGGAGATGGAGCAGAGAAAAATACACATGTTTAGCAAGGAGCTTTTACCAAAAAGGTTGTTTGTATTCACACCCCTGCTTAAACGTTTGGATGGAGAGGAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTAAGATAAAGAAAGCATACTGCCAGCTGGAGTTTGAAGAAATCCAAATATGAGATAGCTAAATACCTCTCTGATATCTTTAACCAATAAAGGCCAGAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAGTTAGAGATTTATTTAAAAATAGGAATTCGATCCAAATGATTAAAAAATGCTGTAGCTGAAGAATCTATAAGATTTTAGAGCCAAATAGAAAGAGATTATAA	Aminoacyl tRNA synthetase clone for the incorporation of p-azido-phenylalanine (p-Az-PheRS(6))	RS
31	ATGGACGAATTTGAATGATAAAGAGAAACACATCTGAAATTTACAGCGAGGAAGGTTAAAGAGAGCTTTAAAAAAGATGAAATAATCTGCTACGATAGGTTTGAACCAAGTGGTAAAAATACATTTAGGCGATTATCTCCAAATAAAAAGATGATTGATTACAAAATGCTGGATTGGATATAATATATTTGGCTGATTTACACGCCATTAAACAGAGAAAGGAGATTGGATGAGATTAGAAAAATAGGAGATTATACAAAAGTTTGAAGCAATGGGTTAAAGGCAAAATATGTTTATGGAAGTTAACTTCCAGCTTGATAGGATTATACACTGAATGCTTATAGATTTGGCTTTAAAAACTACCTTAAAGAGGCAAGAGAGGATATGGAATTTTAGAAGAGAGGATGAAATCCAAAGGTTGCTGAAGTTATCTATCCAAATATGCGAGTTTATCCCTCTTATTAAGAGGCTTGAATTTGCAAGTTGAGGAGATGGAGCAGAGAAAAATACACATGTTTAGCAAGGAGCTTTTACCAAAAAGGTTGTTTGTATTCACACCCCTGCTTAAACGTTTGGATGGAGAGGAGATGAGTTCTTCAAAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGCTAAGATAAAGAAAGCATACTGCCAGCTGGAGTTTGAAGAAATCCAAATATGAGATAGCTAAATACCTCTCTGATATCTTTAACCAATAAAGGCCAGAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAGTTAGAGATTTATTTAAAAATAGGAATTCGATCCAAATGATTAAAAAATGCTGTAGCTGAAGAATCTATAAGATTTTAGAGCCAAATAGAAAGAGATTATAA	Aminoacyl tRNA synthetase clone for the incorporation of p-azido-phenylalanine (p-Az-PheRS(20))	RS

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
32	GGAAATTTTATAGCTGTGATGACTCTCCAGAGAGATTAGGGCTAAGATAAAGAAAG CAATCTGCCAGCTGGAGTGTGTGAAGGAATCCAAATATGGAGATAGTAAATATCT TCCTTGAATATCTTTTACCAATAAAGGCGAGAAATTTGTGGAGAGATTGACAG TTAAATAGCTATGAGAGTTAGAGAGTTTATTAATAAAGGAATTCATCCAAATGG ATTTAAATAATGCTGTAGCTGAAGAACTTATAAAGATTATAGAGCCAAATGAAAGA GATTA ATGACGGAATTTGAAATGATAAAGAGAAACACATCTGAAATTTATCAGCGAGGAAGAG TTAAGAGAGGTTTAAATAAAGATGAAATAATCTGCTACTATAGTTTGAACCAAGT GGTAAATATACATTTAGGGCTTTATCTCCAAATAAAGAGATGATGATTTACAAAT GCTGGTTTGAATTAATATATTTGATGATTTACACCTTTATTTAACCGGAA GGAGAGTTGGATCAGATTAGAAATAGGAGATTATACAAATAAGTTTGAAGCA ATGGGGTTAAAGGCAAAATATGTTTATGGAAGTTCGTTCAGCTTGATTAAGGATTAT AGACTGAATGCTATAGTATGCTTTAAATACCTGCTTAAAGAGCAAGAAAGCT ATGGAATCTATAGAGAGAGGATGAAATCCAAAGGTTGCTGAAGTTATCTATCCAA TAATGCAAGGTTATCCACTGCAITATCAGGGGCTTGATGTGCAAGTTGAGGAGATGG ACGAGGAAATATACACTGTAGAGAGGAGGTTTACCAATAAAGGTTGTTTGA TTTCAACCCCTGCTTACCGGTTTGGATGGAGAGGAAAGATGAGTTCTTCAAAAG GGATTTTATAGCTTTGATGACTCTCCAGAGAGATAGGGCTAAGATTAAGAAAG CATATCTGCCAGCTGGAGTTTGAAGGAATCTCAATAAGGAGATACATAATACT TCCTTGATATCTTTTACCTTAAAGAGCCAGAAATTTGCTGAGATTGACAG TTAAAGCTATGAGAGTTAGAGAGTTTATTAATAAAGGAATTCATCCAAATGG ATTTAAATAATGCTGTAGCTGAAGAACTTATAAAGATTATAGAGCCAAATGAAAGA GATTA	Aminocyl tRNA synthetase clone for the incorporation of p- azido-phenylalanine (p-Az-PheRS(24))	RS
33	ATGAGCGAATTCAGGATTAATTCAGGAGAAAGTGCGAGAAAGCGTGGGAGAA GGAGAGAAATTTTGAAGTCCGATCTTAATGAAGAGAGAGTTTCTTCTCTCA CAATCTCCCA TCCCTACTCTTAAGGAAATCTTACAGGAGGTCACAGGAA ACCTTACAAA TTTGGGATGCTCTTCGCCAGA TACTAGAGAA TGAAGGGCTA CAACGTTCTC TTTCCCTCCG GCTTTCAATGT TACGGGACCC CCAATCATTTG GCTTTGCGGA GCTCTATGCT AAGAGGAGCG AGAGCAATAG AGAGGTTTAC ACCAATAACC ATGACGTTCC GCTGGAGGAC TTGCTTCAGC TCACAACTCC AGCAAAATCC GTTGAATACT TCTCAAGGGA GCGCTCGAG GCTTTGAAGA GCATAGGCTA CTCATTGAC TGGAGAGGG TTTTCAACC AACCGATGA GAGTATCAG GATTCCTCGA CTGGCAGTAC TGGAGGTTCA AGAGGCTTGG CCTGATTTGG AAGGGCACCC ACCCGCTCAG ATACTGCCCC CAGACCCAGA ATCTGTTTGA AGACCAAGAC CTCTCTCTG GGAAGGAGCC AGTATTTGTT GAATTTACCG TTATTAAGCT CAGGCTTGA GATGAGAGCT TCATTTTCC CTGTGCAACT CTCCTCTCCG AAACCTGTGT TGGGTCACG AACATCTGGG TAAAGCCGAC AACCTACGTA ATTGCCOAGG TGGATGGGA AAATGTGTTT GTGACAAAG AGCTTTAGA GAAGCTTACC TACAGGAGA AAAAATTCAG GCTCCTGGAG GAGGTTGATG GCTCGAGTT CTTCGCAAG TACGTATAG TCCCGCTGGT AAACAGAAA GTGCCAATTC TGCTCGCAGA GTTGTGTGAC ACCGAGAAC CAACAGAGAT TGTGTGAG GCTTCGKAC AGGCTCTTT TGACCTGGCT GCCATTGAGG ACTTGAAGAG AGACAGAGAA ACCTGTGCGA AGTACGAGAT TGACAAAAGC GTTGTAGAGA GCATAAAGCC ATAGTCTCTG ATTAAAGAGC ACATTAAGG TTTCTCTGCT GAGAGCTTAA TAGAGAGCT TGBHSHAG AGCGBHAGG ACAGAGAGCT GCTGATAGG CCACACAGA CCCTCTACAA GAGGAGATAC CACACGGGAA TCACTGTGGA CAACAGATG AACTATGCTG GAATGAAAGT TCTGAGGCG AAGAGAGAG TTCTAGAGGA TTTGTGTAG CTGTGCTTGG GGAATTTTCT CAGCACTTC AGCGAGAGC CCGTAAATCTG CAGTCTCGGA ACCAAGTCCG TTGTTAAGGT TGTTAGGGAC CAGTGTCTCC TGAATCTACT CACAGAGAG TGGAGAGAGA AGTCTCTGAA TCACTTGAA AAGAGCGGA TCTCTCCGTA CACATCAGG GAGGATCTCA GGAAACAAGT TGAATGGCTC AGGAGCAAGG CTGTGTCCAG AAGGAAAGGG CTTGGACAGA GAATTCCTGG GATTAAGAGG TGGCTCATCG AGAGCTTTT AGCTTACAGA ATCTACATGG CTAATTAAGT CTTTCCAG TACATCAAG CAGGATTTGCT CAGCGCCGAG AACATGACTC CCAAGTTCTT CGACTACGPG CTGTGGGCA AAGGTGAGGT TGGAAAGTT GCGAAGCTT CAAACTCAG CGTGAAGTTA ATCCAGAGA TCAGGAGCA CTTCGAGTAC TGCTATCCG TTCACCTAG AAGCAATGCG AAGCACTTGG TTGCACCA CTTGCTCTTC TAACCTTCCC ACCACGTCGC CATTTTCCCG CCAGATAAGT GCGGAGGGC GAATTCGCTA AACGATACG TCAGGCTTGA GGGCAAGAG ATGAGCAGA GCCAAGGGCT CTGTCTAAGC ATCAAGAGGG CGGTCTAGCA GATGATGCG GATGTGACGA GGCTCTACAT CTTCCAGCTT TCCAAACCA CACAGGATGC GGACTTGAAG AGCAGAGAGG TTGAAGGCT TCCAAACCA CTCAGGAGGT TCTACAACT CTGAGAGAG AACTGCTTA AAGGATGAG AGAGTAAK ACCTCGACCC GCTGGCTTGT GAGCAGGATG CAGAGGGCAA TAAAGGAAGT GAGGAGAGCT ATGACAACC TGCAGACGAG GAGGCGCTGT AATCGCTCT TCTTGAAGCT CAGGAAACG CAGATAGGT ATCTGAGAG AGGAGTTGAG AACTCTGCTA TAATCTGGA OXACTGGATC AAGCTCTCG CCCCCTTTCC TCCGCAATTT TGCAGAGAGC TGTGGCACTT GAACATGAC AGCTACGTCA GCTCTGAAAG CTACCCAGAA TACGCGAAA CAGAGTTTGA CAGAGAGGGC CGAGGATTTG AGAATACCTT CCGAAACCTT GTTACGAGCA TTCAGGAAT CAAGAAGTTT GTTACGATG CGAAGAGAGT TTACATTTCT CCCCCTTGG ACTGGAAGTT TAAGCACGA AGGTCTCTG CTGAAAGCG GATGTTTGG GAGGCGATGA AGCTCTTAT GCGAGACGAG GACTTTTGA AGCTTGGCA AGAAGTGTCA AATTTCTTCA AGAAGATTTT CAAAGACAGA AAGAAGCTGA TGCTAGTTAA GAGATGGGAA GTTCTGACG AGAATCTGAA ATTTATTGAG AAGAGAGCG GACTGAAGGT TATCTCTGAT ACTCMAGG TTCTTGAAGA GAAGAGAGG CAGCAGATT CCGGCAAGCC CCGGATTTAT GTTGTCTAA	Archaeoglobus fulgidus leucyl tRNA- synthetase (AFLRS)	RS
34	GTGGATATTTAAAGAAAATG GGTGTATAGA TGGAGAGATG CTGGCATATT TCAGGCTGAC CCTGATGACA GAGAAAGAT ATCTCTCACA GTGCTTACC	Methanobacterium thermoautotrophicum	RS

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
	CCTACCCAG TGGTGGGAT CACATAGGAC ACGGAGGAC CACACTGTC CCGATGCTC ATCCAGGTT CAAGAGGATG CAGGGGTACA ACCTCTGTT TCCATAGGC TGGCATGCA CAGGGGCCG TGTCTAAGG ATAGCGCGA GGAATCAGAG GAAGGATCCC TGGACCTCA AAATCTACAG GAGGTCCAC AGGATCCCG AGGATGAGCT TGAACGTTT AGTGACCTG AGTACATAGT TGAATATTC AGGAGGAAAT ACCGCTCTGT TATGAGGAT ATGGCTACT CCATCGACTG GAGGCGTGA TTCAAAACCA CGATCCAC CTACAGCAG TTCTATAGT GGCAGATAAG GAAGCTGAGG GACCTTGCC TCGTAAGGA GGGGCCCAT CCGTGAAGT ACTGGCTGA ATGTGAAAC CCGTGGGTG ACCGTGACCT CCGTGAAGT GAGGGGCTG CCGTAAACCA GCTGACCTC CTCAAATCA AACTTGGAGA CTCATACCTG GTGCGAGCCA CPTTCAGGC CGAGCAATC TATGGGCCA CCAACCTCTG GCTGAACCT GATGAGGAT ATCTGAGCT TGAACACCT GTGAGGATG GGTAAATAG CAGCTTCTC GTGATATTC TTTCACACCA GAAACTGGAC CTCAGGTTT CCGGTGACGT CAACCCCGG GAGCTGATAG GGTATGCTGT GAGAAATCT GTACGGGCC AGGAGACCC CAGACTCCCG CTTCTCTTG TTAGACCTGA ATATGCGCA GGTGTGTGT TCTCTGTCCC TGACATGCC CTTGAGACT TCATAGCCCT TGAGAGCTC AGGAGAGACC ATGAACCTCT TGAAGGTAC GGTCTTGAGG ATGTGTTTC TGAATATGAG CCGTGAAGT TCATAGCAG GATAGGCTAC CCGTATTC CCGGCCCGA GGTATAGAG AAATTTGGT TCAGAAACCA GGAGGACCCC CGCTTGAGG ATGCCACCG GAGCTATAC AAGATCGAGC ATGCGAGGAG TGTATGAGG AGCCATCTC CTGTCTATG TGTATGAGG GTCTCTGAG CCGTGAAGT CATTCTGTT GAGTGAAGG ATCGGGGCT TGAGATGAG ATGTATGAT TCGCTGAGG ACCTGTATTA TCGCGCTGG GTGCGAGGT CPTTGAGG GTCATGAGG ACCAGTGGT CATGAGTAC TGTATGAG CCGTGAAGG CTCGCCAC AGGTGCTGT ATGCTATGA GATATACCC GAGGAGTCC GGGCCAACT TGAATACTAC ATGACTGGC TCAATGACT GGCATGTTCA AGGAGGATG GCTTGAAGC AAGCTGTCC TGTATGAG GGTGATCAT CCAACCTCTC AGGAGTCA CAATCTACT GGCATATTAC ACCATGACAC ACCGCTCAG GAGATGGAT GCGGGGAGA TGAGCATGA GTTCTTTGAT GGCATATCC TAGATGATC AGGAACCTT GAGATCTCA GAGAGGATC CCGTATGAG TACCCCTTG ACTGAGCTC CTTCTCAAG GACCTGATG GCAATCAGT GACATTCAT ATATCCACC ACTCAGCAT ATTCTCTGAG TCAGGCTGC CCGGGGGGC TTGTGCTTT GGTATGGCC TTCTTGAGG CAGCAAGATG TCATCTCCA AGGCAAGCT CATATCTG AGGAGCCA TCGAGAGCA CGTGCAGAG GTGTGCGGC TCTCTCAT GTCTCAGCA GAGCATGGC AGGACTTTGA CTGAGGAG AGTGAAGTCA TCGGAGCCG CAGGAGGAT GAATGTTCA GGGAAATCG AGAGAGGCT TCAGGTATCC TGAATGATG GCGCTGCTC AGTGAATTA CTCGAGCTGA ACCTGAAGC TTCAATGAA GTTGATGAT GGTCTGCTG AACAGAGGA TACGTGAAG CACAAGGCC CTTGAATCAT TCGAGCAAG AAAGAGGCT CAGAGGCC TCTATCTCT TAAAGAGAT GTGACCAT ACCTTAAGC TGTGAGGT AGAGTTGATG ATGAGGTTAA ATCTGCTCT GCAAGGTC TGCAGGCTC GATAAGGCTC ATGGCTCAT TCATACCTTA CACTCTGAG GAGGTGGG AGAGGTATG TGTGAGGCT TTTGTAGAG AACCTCATG CCGTACTTC TCAGATGATG CAGAGAGCAG GATGTGAG GTTGCAGAG AGATGCTCA GAATACCTT AGAGACTTC AGGAATCAT GAGATCTCT GATCTACT CAGAGAGGT CCATATATC ACTGACCA ATGCAATG GATCTCTT AGGCTGAG CAGAGTAGG AAACTAGAT ATGGGCTCA TAATGGGAG GGTTCAGCT GAGGCTATC ATGATAACAT GAGGAGGCT GCTGAATTT TAGGAGGAT CATCAGGAC CTGTTAAAT CAGGCTGAG GTGATGAG GGTGAGGCT TACTGAGG TCGCTCTGAT TACATTAAT CAGAGTGG AGCCAGGCT GTGATACCA GCAACACGA CTATGACCT GAAACAGG CTTGTGATGC CTTCCCTG AAGCAGCCA TATACCTGA AGA	leucyl tRNA-synthase (MtLRS)	
35	MDEFEMIKRNTSEIISEELREVLKKDEKSAIGFEPGSKIHGHYLTQKKMIDLQ AGFDIILLADLHAYLNQKGLDEIRKIGDYNKVFEMGLKARYVYGSTPQLDKDY TLNVYRLAKTTLKRAHSMELLAREDENPKVAEVIPTMQVNPZHNGVDVAVGGM EQRKIHMLARELLPKKVVCTHNPVLTGLDGEKMSSSKGNFTAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPFGGLTVSSYEELSLFKNKELHFM DLKNAVEELIKLEPIERRL	mutant TyrRS (LWJ16)	RS
36	MDEFEMIKRNTSEIISEELREVLKKDEKSAIGFEPGSKIHGHYLTQKKMIDLQ AGFDIILLADLHAYLNQKGLDEIRKIGDYNKVFEMGLKARYVYGSTPQLDKDY TLNVYRLAKTTLKRAHSMELLAREDENPKVAEVIPTMQVNPZHNGVDVAVGGM EQRKIHMLARELLPKKVVCTHNPVLTGLDGEKMSSSKGNFTAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPFGGLTVSSYEELSLFKNKELHFM DLKNAVEELIKLEPIERRL	TyrRS (SS12)	RS
37	MDEFEMIKRNTSEIISEELREVLKKDEKSAIGFEPGSKIHGHYLTQKKMIDLQ AGFDIILLADLHAYLNQKGLDEIRKIGDYNKVFEMGLKARYVYGSTPQLDKDY TLNVYRLAKTTLKRAHSMELLAREDENPKVAEVIPTMQVNPZHNGVDVAVGGM EQRKIHMLARELLPKKVVCTHNPVLTGLDGEKMSSSKGNFTAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPFGGLTVSSYEELSLFKNKELHFM DLKNAVEELIKLEPIERRL	p-iPr-PheRS	RS
38	MDEFEMIKRNTSEIISEELREVLKKDEKSAIGFEPGSKIHGHYLTQKKMIDLQ AGFDIILLADLHAYLNQKGLDEIRKIGDYNKVFEMGLKARYVYGSTPQLDKDY TLNVYRLAKTTLKRAHSMELLAREDENPKVAEVIPTMQVNPZHNGVDVAVGGM EQRKIHMLARELLPKKVVCTHNPVLTGLDGEKMSSSKGNFTAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPFGGLTVSSYEELSLFKNKELHFM DLKNAVEELIKLEPIERRL	p-NH ₂ -PheRS(1)	RS
39	MDEFEMIKRNTSEIISEELREVLKKDEKSAIGFEPGSKIHGHYLTQKKMIDLQ AGFDIILLADLHAYLNQKGLDEIRKIGDYNKVFEMGLKARYVYGSTPQLDKDY TLNVYRLAKTTLKRAHSMELLAREDENPKVAEVIPTMQVNPZHNGVDVAVGGM EQRKIHMLARELLPKKVVCTHNPVLTGLDGEKMSSSKGNFTAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPFGGLTVSSYEELSLFKNKELHFM DLKNAVEELIKLEPIERRL	p-NH ₂ -PheRS(2)	RS

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
40	EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	p-NH ₂ -PheRS(3a)	RS
41	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	p-NH ₂ -PheRS(3b)	RS
42	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	O-Allyl-TyrRS(1)	RS
43	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	O-Allyl-TyrRS(3)	RS
44	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	O-Allyl-TyrRS(4)	RS
45	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	p-Br-PheRS	RS
46	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	p-Az-PheRS(1)	RS
47	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	p-Az-PheRS(3)	RS
48	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	p-Az-PheRS(5)	RS
49	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	Mutant synthetase to incorporate m-acyl phenylalanine into proteins (Ketone 3-4)	RS
50	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	Mutant synthetase to incorporate m-acyl phenylalanine into proteins (Ketone 3-7)	RS
51	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFNKKELEHFM DLKNAVAEELIKILEPIKRL	Mutant synthetase to incorporate m-acyl phenylalanine into proteins (Ketone 4-1)	RS
52	MDPEFMKRNPTSEIISEELREVLKKDEKSAIIGFEPGSKTHLGHYLQIKKMDLQNG AGFDITILLADLHAYLNQKGELEIRKIGDYNKVFAMGLKAKYVYGSPFQLDKDY TLNVTYRLAKTTTKRARRSMELLAREDENPKVAEVIYPIQVNTYHYGVDVAVGGM EQKTHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK	Mutant synthetase to incorporate m-acyl phenylalanine into	RS

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
	AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	proteins (Ketone 5-4)	
53	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	Mutant synthetase to incorporate m-acyl phenylalanine into proteins (Ketone 6-8)	RS
54	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	Mutant synthetase to incorporate m-methoxy phenylalanine into proteins (Ome 1-6)	RS
55	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	Mutant synthetase to incorporate m-methoxy phenylalanine into proteins (Ome 1-8)	RS
56	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	Mutant synthetase to incorporate m-methoxy phenylalanine into proteins (Ome 2-7)	RS
57	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	Mutant synthetase to incorporate m-methoxy phenylalanine into proteins	RS
58	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	Mutant synthetase to incorporate m-methoxy phenylalanine into proteins	RS
59	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	Mutant synthetase to incorporate p-O-allyl tyrosine into proteins	RS
60	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	Allyl Aminoacyl tRNA synthetase for the incorporation of p-benzoyl-L-phenylalanine	RS
61	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	p-BpaRS(H6) Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine	RS
62	MDEFEMIKRNTSIISEELREVLKKDEKSAITGFPSGKIHLGHYLQTKKMDLQNG AGFDIILLADLHAYLNQKGEDEIRKIGDYNKKVFEAMGLKAKYVYGSFQLDKDY TLNVYRLAKTKTKRARRSMELIAREDENPKVAEVIYPIQVNDIHYLGVDVAVGGM EQRKIMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMETAKYFLEYPLTIKRPKFGGDLTVNSYEELSLFKKELHPM DLKNAVAEELIKILEPIKRL#	p-Az-PheRS(3) Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine	RS

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
63	MDPEFMKNTSEIISSEELREVLKKDEKSAITGEPSPCKIHLGHYLQIKRMIDLQN AGFDITLLADLHAYLNQKGLDEIRKIDYNNKVPFAMGLAKYVYSSFDLQDY TIANVRLAKTTLKARRSMELLAREDDPKVAEVIYPIHONVPVHYQGVDAVAGSM EQRKTHMLARELLPKRVCTHMPVLTGLDGEKMSSEKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRFKFGDLTVNSEELESLFANKEIHPM DLKNVABEELKILEPFRKL	<i>p</i> -Az-PheRS(6) Aminoacyl tRNA synthetase for the incorporation of <i>p</i> - azido-phenylalanine	RS
64	MDPEFMKNTSEIISSEELREVLKKDEKSAITGEPSPCKIHLGHYLQIKRMIDLQN AGFDITLLADLHAYLNQKGLDEIRKIDYNNKVPFAMGLAKYVYSSFDLQDY TIANVRLAKTTLKARRSMELLAREDDPKVAEVIYPIHONVPVHYQGVDAVAGSM EQRKTHMLARELLPKRVCTHMPVLTGLDGEKMSSEKGNFIAVDDSPETRAKIKK AYCPAGVVEGNPIMEIAKYFLEYPLTIKRFKFGDLTVNSEELESLFANKEIHPM DLKNVABEELKILEPFRKL	<i>p</i> -Az-PheRS(20) Aminoacyl tRNA synthetase for the incorporation of <i>p</i> - azido-phenylalanine	RS
65	MSDFRIIEK WQKMSKDR FESDPNEK FPLTTPYPL NGMLHAGHTA TPTIGDAP YHMMGVWL FPLSPVITV PIIQIAGLIA KREERTLEY TKYHDVPLED LQLTTFEKI VEYFSREALQ ALKISYISID WRRVPTTDE EYQRTIEWY WKLEIGLIV KOTHPVRCF HDQNFVEDHD LLAGEEATIV EPTVTKFRIE DQDILFCAT LARSPFQFT NIKWKFVTV TAVUGEDQW VSKZAYEKIT YTEKKVRLLE EVDASQFGK YVIVPLVNRX VPILPAEFVD TINATGVVMS VPAAAPEDLA AIEDLKRDEE TLAKYGIKDS VVESIKPIVL IKTDLDVPA EKILAEIGVE SQKQELLOK ATTKLYAKVE HTGIMLDNM NAGMMVSEA KERVDDELVK LGIGDVVEF SEKPVICROG TKCVVKVVRD QWFLAYSNRE WKEKULMHLE JORLIIPVYK EPRANKIWL RDKACARRRG LQTKIPMDKE WLIESLSST ITHAYTILAK YNAGLLKAE NTPPEFLDVI LQKGEVOY ARSKLSGVL IQIIRDDFY WYFOLDSRG KULVANHLLF YLPHVAIIP PDKWPAIAV NGVSLGKK MSKSGPLIT MKRAVCQYGA DYTHYILHA AEYDSADWK SREYGLASH LAREFNLVRE NYLKEVGLIT TLDMLVSM ORALIEPERA NMLQTHAV NAFPELMD VNWTLRGGE NLATLLDDVI KLLAPFAPIT CBRMLHMD SYVLESYPE YDETRVDEA ERIEYLRHL VEDIQETRF VSDAKETVTA PAEDMKVAA KVAARESDVG BAMKQMDR ELKLGKEVS NFVKYIFOR KLMLVKEW VLQQLKPIE NETQLKVID TORVPEEKRR QAVPGKPIY VA*	Archaeoglobus fulgidus leucyl tRNA- synthetase (AFLRS)	RS
66	VDIERKWRDR WRDAGIQAD PDREKIFLT VAYPYPSGAM HIGHRTYTV PDVIRFMR QGINLPPA MRYTAPVIG IANKIQHOP MTKITRSDH RVPKDLSEF SDPEYVEYF SREYRSVNEED MGYSIDWRRE FKTTDPTYSR FVQWQIRKL DLGLVRGAH PVKVCPCBN PVGHDHLEG EGVAINGLTL LKFILGDSYL VAAITRFRPI YGATNLWLNF DEDIVRVETG SEEWI BRAA VNLSHQGLD LKVSQVNRG DLICWVRE VTCHEPILP ASVQVEYAT GVFVSFAHA PADFTALBDL RTHELLERY GLEDVVADIE PUNVIAVDGY GEPPAEVIE KFGVRQBDP RLEDATGLY KIEHARGVMS SHIPVYGGK VSEARVID ELKQGLADE MYEPASRPI CCCKGCVTV VKDCNGMKY SDDAKDLAH RCLDGMKII P EKVRAFETV IDWLNDWACS RRIIGLTRLP WDERWIEPL TDSPTVMAYV TIAHRLREMD AGEHDEFFD AIFLDUSGTF DLAEERFVW YPLDRLGAK DLIGNLITH IFHSAIPE SOMPRGIVP GMGLLEGKMM SSKGNVILL EDAIEKHAD VVRLFMSSA EPWQDFWRE SEVITRRRI EWFREFGERV SGLDGRVPL SEVTPAESPS PIGRMMQGL NQRIRATRA LESPTKRAV QEALYLLSD VNHVLAKEG RVQDEEVSIL ANLHAHRL MAFPIFYTAZ EMERYGQEG FVAAPWPDF SDAERSRDVQ VAEDMVQTV RDIQIMKIL GSTPERVHY TSPKRWMDVL RVAAEVGKLD MSINGRVA EGHDMKREY AEPVRLILD LKSEVTVID SYSVLMDASD YIESEVGA VTHGFVDYD ENKAVWAFV KALYLE*	Methanobacterium thermoautotrophicum leucyl tRNA-synthetase (MILRS)	RS
67	GAMTTCACAC ACAGGAACAC GCTATGCCCA CGCTCTGAT CGACAACAC GACTGCTTCA CCGAGAACAT GTTCCAGPAC ATCCGCGAGG CCACCGGCA GCCCGCGTC GTGCCAACG ACCCGGAGTC GTCCGCGTC CCCCGGAG ACTTCGACGC GATCTGCTGG TCCCGGCGCC CGCGACGCC CGACCGGAA CGGACCTTCG GATTCAGCCG CCGCGGATC ACOCAGAGCG GCTGCGCGT CTCTGCGTC TCGCTCGGCG ACCAGGACAT CGCCGAGTC TCGCGGAA CCATGACGG CGGCTCTCC GAGGTGCGGC ACACCGGCA GACGCTCTC CGGCGCTCC CCGCGCGTT CACCGCGGT CGCTACACT CCGTGGCGC CACCGACCT CCGCACMG CCGACCGCT CGCTGAGG GAGCGGCG TCCTCATGG CCGCGGAC CCGGAGAG CGCTGATGG CTCCAGTTC CCACCGAGT CCATCGGAC GCACTTCGCG CGGAGATCA TGCCCAACTT CCCGACCTC GCTTCGCGC ACCACCGGC ACOTCGGAC GCGGCGACT GGGCTTGA ACTCACGTC CCGCGGTC ACCTGCTCC GACACCGGA GAGTACGCC CGCTGCGTC CCGCGCGAG GCGCGACGT TCTGCTGGA CAGGACCTC GTCTCGAAG GCGCTCGCT GTTCTCTCTT CTGCGGAGC ACCGGAGTC GTCTCGGAG TACTGCACT ATXKCTGCG GCGCGGCTC GTCTCGCTCC CGGCTCGCA CGGACACAG ACCCGGAGC GCGGACCTC CTTCAGCTAC CTGAGGAGC AGCTCGAAC GCGCGCGGT CCGCTCGCTC CGACCTGAC CTTCGAGTTC AACTCGGCT AGGCGGCTA CCGTGGTAC GAGTGAAGG CGGAGACAC CGGCGACCC GCAGTACCG CCGCGACCC CGACCGCGG TTCTCTCTTC CGACCGGCG CATCGCGCT GACACGAGG AAGGCTGTC CTACTGCTG GCGCTGAC CCGCGGAGCA CGACGAGCG GCGCGGCTT GCGTCCGGA GACGCGGAG ACCTCACCG GCGTGGCGGT CCGCTCGCG CGGAGGCGA CCGCGCGAT GTCTCTCGG GTCCCGGAG CGGCGCGCG CTTCGCGGCC CTGCTCGCG CACGCGAGA CAGGAGCGC	(plac-papabc)	Plasmid

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
1	TCGGCGCTCC GCAACGGCGA GTCGTACGAG ATCTGCGTGA CCAACAAGGT		
2	CACGCGCGCC ACCGAGGCGA GGGCGCTGCG GCTGTACTCC GCGCTGCGCC		
3	GCATCGAGCC CGCGCGCTCT GGGCGCTGCG TGGAGTCTCC GGAAGTGTGG		
4	GTGCTTAGCG CCTCGCCCGA GCGGTCTCTC AGGATCGGCG CGACGCGCGG		
5	CGTGAATCC AAGCCCATCA AGGGGACCGG CCGCGCGCGG GCACCGCGCG		
6	AGGAGAACCA GCGCGCTCGG GCGGACCTGG CCGCGCGGGA GAAAGGACCG		
7	GCAGAACCT TATGATATGG GAGCTGTGTC GCAACGAGCC TCAACAGCGT		
8	CTGCGCGATC GGTCTCGTCC AGTGGCGCGG GCTCTTCTAG GTGGGAGAGCC		
9	TCGGCGCGGT GCACAGCTG GTGTGACACA TCCGCGGAGC GCTGCGCGCC		
10	GCATCGAGCA CCGCGCTCTG CTGAGCGCTC GCGTCTCGCG GCGGCTCAT		
11	GACCGCGCGG CCGAAGAAGC GAGCCATGGA GATCATCGAC GCGCTGGAGG		
12	AAGGCGCGCG GGGCGCTCTA CCGCGCGCGG TCGGATGGTT GCGCTTCAAG		
13	GCGCGCGCGG ACCTGAGCAT GTTATCTCGG AGCATCTGCG TGGCGGAGCG		
14	CCGCGCGGAG TTGCGCTCGG CCGCGCGGAT CGTGTCTCTC TCCGAGCAGG		
15	AGGAGGATTT CAGGAGAGCC GTGGTCAAGG CCGCGCGCAT GGTTCACGCG		
16	CTGAGAGGCA GCGGAGGCG GGGCGGAGCA TGCACCAAAC AAGGAGCTTA		
17	GATCATGACC GAGCAGAACG AGCTGACAGT TCGCGCTCGG CCGGAGCTTC		
18	GAGCGCGTCC ACGGAGAGCT TCTGGACACG GTGCGCGCGC GCATCGAGCT		
19	CGGTGTCCCG ATGCGCGGCT ACAAGTCCCG GCACGCGCTC CGATGATGCG		
20	AGCTCGCGCG GGTCACTCTG GTCAAGGACA GCGCGCGCGC CTACGCGCGC		
21	GACCAAGGCG TCGAGAGATC GTTCTGTGTT AACCTCTACG ACGTGAATCAT		
22	CAGGAGATGT TCGCGCGTGG AGGACCTGGT GATGAGCTCG TCAATGATCTA		
23	AGGAGGTTGT ATGAGTGGCT TCCCGCGGAG CGTCTGTGTC GGGGCGAGCG		
24	GAGCGGTGGG CCGCATGTTT GCGCGGCTGG TCGCGGAGCG GGGCAGCGCG		
25	ACGCTCGTGG TCGACCTCGT ACCGCGCGCG GAGCGGCGCG AGCTTTCCTT		
26	GGTGGGAGC GTCACTCGCG CCGCGCGCGA GCTCTGCGCG GCGCTTCCCG		
27	ACCGGAGCGT CGTCTCTGCT GCGGTACACG AGGACGTGGC CCGTCAAGCG		
28	GTGGGCGCGG TGACCGCGCT CATGCGAGCG GCGCGCGTGC TCGCGGAGCA		
29	CGTGTCTGTC CGGAGCGGCA TGCGCGCGCA GCTCTGCGCG CAGGCGCGCG		
30	GCGTCCAGCA CGTGGCGCTC AACCGGATGT TCGCGCGCGC CGCGCGCATG		
31	ACCGCGCGCG CGGTGCGCGC CGTGTCTCAC AGGAGCGCGC CCGCGCGTCA		
32	GACCTCTGCT GCGCGCTGCT AGGCGCGCGG CGCGAGCGCG GTACGCGCTA		
33	CGCGCGAGCA GACGAGCGCG AGGAGCGCGG CGACCGAGGC CCGTCAAGCG		
34	GCGGTGATCC TCTCTCTGCG GCTCGCGCTC GCGCGCGCTC GCGTCAAGCG		
35	CGCGCGCTCT GCGCGCGCGC GACCGCGCGC GCGCGCGCTC GCGTCAAGCG		
36	TCCTGTGCTG TCTGTCTGCG GCGCGCGCGC AGGTGTACCG GCGCATCTAG		
37	CGGTCAAGCC CCGCGCGCGC GTCTCGCGCG CCGCGCGCTC CCGAGCGCGT		
38	GCGCTCTCTC GCGCGCGCTA TCGCGGAGCA CCGCGAGCGC GCGGAGAGCC		
39	CGGAGCGCGC CGAGAGAGCT GAGCGGAGCG AGCAACTCGG CCGTCCCGCG		
40	GGAINGAGCG GCGCGCGGAA CCGCGAGCGC GTCTTGGAGG AACCTCGCGC		
41	GCTCATGAGA CGGAGAGCTG CCGCGCGCGA GAGCGAGTGC CAGGAGCTGT		
42	TCCGACCGCT CCGCGCGAGC GAGCGAGAG GCGGAGAGCA CGGAGAGT		
43	TAGGTGACAC TATAGGAGAT CTCTAGCGCG GAGCATCTGT GCGCGCGCAT		
44	ACCGGCGCCA CAGGTGCGGT TCGTGGCGCG TATATCGCGG ACATCAGCGA		
45	TGGGAGAGAT CGGCTCTGCG ACTTGGGCTT CATAGCGCTT TGTTCGCGCG		
46	TGGGTATGAT GCGAGCGCGC GTGGCGCGCG GACTGTGGGG GCGCATCTCC		
47	TTGCTATGAC CATCTCTTGC GCGCGCGGCT CTCAAGCGCC TCAACCTACT		
48	ACTCGGCTGC TTCTTAATGC AGGAGTCCCA TAAAGGAGAG CCGTCAAGCG		
49	TGCCCTTGGG AGCTCTGAG CCGTCAAGCT CCGTGGGCTT GCGCGCGGCG		
50	ATGACTATCG TCGCGCGGCT TATGACTGTC TTCTTATATG TCGAAGCTGT		
51	AGGAGAGGTT CCGCGAGCGC TCTGGGTGAT TTCTGGGAG GAGCGCTTTC		
52	GCTGAGAGCG GAGGATGATC GCGTCTGTC TTGCGGTATG GCGCATCTG		
53	CAGCGCGTCC CTCAAGCGCT CGTCACTGCT CCGCGAGCCA AACGPTTGGG		
54	CGGAGAGGAG GCGATTTATG CCGGATGCGG GCGCGAGCG CCGGCTTACG		
55	TCTTCTGGCG GTTGGCGAGC CCGGCTGAGA TGCGCTTCCG CATTTATAT		
56	CTTCTCGCTT CCGCGCGCAT CGGATGCGCG GCGTTCAGG CCGTCTGTC		
57	CAGGAGGATG GATGAGGAGC ATGAGGAGCA GCTTCAAGGA TCGCTCGCGG		
58	CTCTTACGAG CTTAATCTCG ATGAGTGGAG CGCTGATGCT CAGCGGAGT		
59	TATGCGCGCG GCGCGAGCAC ATGAGAGCGG TTGGCATGGA TTGTAAGCGC		
60	CGCGCTATAC CTGTCTGCGC TCGCGCGGTT GCGTCTGCGT GCATGAGAGC		
61	GGGCGACCTC GAGCTGAATG GAGCGCGCGC GCGATCTCTT AACGATTTCA		
62	CGATCTGAG AATGCGAGCC AGTCAATCTT TCGCGAGGAG TGTGAATCGG		
63	CAAGCAAGCC CTGGGAGAGA CATATCTCAT GCGTCTGCGA TCTCCAGCAG		
64	CGCGAGCGCG GCGATCTGCG GAGCGGTTGG GTCTTGGCCA CCGGTGCGCA		
65	TGATCTGCTT CCGTCTGCTG AGGAGCGCGG TACGCTTGGG GCGGTGCTCT		
66	ACTGTTTATG AGAATGAATC ACCGATACCG GAGCGAGCGT GAGCGAGCTG		
67	CTGTGCGAAA ACGTCTGCGA CCGTGAAGCA AACATGAGTG GCTTCTGCGT		
68	TGCTTGTTCG GTAAAGCTCG GAAAGCGGGA AGTCTCTTAC GTGTGCTGTA		
69	AGTTGCGCGC AACAGAGAGT GGAACCAAGC GTGTATACCA CGATATATG		
70	ACTGAGAGTC AACGCGATGA GCGGCGCTCAT TTCTTATCTT GAGTTACAAC		
71	AGTCCGAGCC GTTGGCGGTA GCTACTGAGC TATGCGGCTG CACTAGCGCT		
72	GCGTCAAGAT GCTCTGATCC AAGCGAAACT GCGAATAATAT CTGCTGGCAG		
73	CGGAGGTGAG CCGCGCTGAC CATTTATGTC CGGATCTGCA TCGGAGGATG		
74	CTGTGCGCTA CCGTGTGAAA CAGCTACATC TGTATATAGC AAGCGCTGAG		
75	ATGAGCGCTG AGTATTTTCT CTCTGCTGCG GCGCTATCTCC TTGTGCGAGC		
76	TTGCCACGCT CAAGGCGGTT TGAAGTCCAA CCGTACGAAA AGTATCGGTA		
77	AGAGGAAATAT TATCGTCTGA AAAATCGATT AGTACAGAG AAGTCCGCTT		
78	AAGTCCGATAT TTTCGATATA AAGAGACGCG TTGTATGAGC GTTCTTCCAT		
79	GTACATTAATG TTTCGATATA TCGAGAGGTT ACTTCACTAA CGTACGCTTT		
80	CGTCTATATG GCGTCTAGAC CCGATATCTT TAATGCTCTT TATTTGCTGG		
81	GCTTATGAGA TCGCGCGCAC ACGTATATAT AAGCTTTTCT CCGCGAGAGA		
82	TGAGCGGACA GCGTCTCTAG TCTATGATGT TATTTCTTAA TAATCATCCA		

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
1	GGTATTCTCT TTATCAACAT ACGTAGTGG AGTGCCACC TTAACGCAGG		
2	GCTTTCGCTC ACAGCGCGAT ATGTGAGCCA GCGGGGCTTT CTTTTGGCAG		
3	ACGGCTTCCA TCCCTCGCAT TCGAGCAATC TGGCTTTCAG CCGCATTCAT		
4	AAACCCAGTA AATGCCGTCA CCGAGGAAGC CAGAGCCGAG AATATCGTCA		
5	GTACAGAGTA AATCGCGGAT TTCCACGTAT AGCTGACAT CTACAGAGCC		
6	ATTTCATGGA TCATCGCTT CCGCTATCG CAGAGCTGAT TCAGAGCTTC		
7	TCTCGCGGT TCTGCTGTG CTAATCGGG TTGTTCAGT CTTTCTCAA		
8	CCTGAGTGG CCGGGAATC ACCGATTTCC TGACGGTGTG AGTCATATTA		
9	CCGAGACGTC TGTCCTGAGT AGCAATGACC CTCTCAGCG TTCTACTTTG		
10	CTCTCTAAT TGTGATGAG GGGCTGAAA CTGTCTGTG AGAGAGATTA		
11	CACGCTTTTC CAGCGCTGA TGATGCCGA TAAGGCGGSC AATTGTGTTA		
12	ATTTCGCTGC TCATACAAA TCCTGCTAT CTGAGAAATG ACCAGCCTTT		
13	ATCGGCTTC TGTCGTAAT GTTCGGGAG TCGCTCTCT TCTTCTCTCT		
14	GCTGAGCTG TTTTTCGGC AGACGTTGCG GCTCTCTCT CTTTTCATC		
15	TCCCTGATGA TCCCTGTGAA CTCGGCATC GCATCGTTAA CAAAGGACTG		
16	AAGATGATT TCTTCTCTGA TACCTCTAT GGCATACATG ACCAGTGGT		
17	TCAGCTTGT AGGCTCTTT? TCAAAATCAA ACCTTCTGCC GSAATGGGAT		
18	TCCTGCTCAG GCTCTGACT? CAGCTCTGT TTTAGCTCA GAGTATCCCT		
19	CTGCTGAGG GCTCTCGGTA AGAGGTAAT CAGCTCAAT ACCGTGTCAA		
20	GTATCATCAG GACTGCTGC ACCTGCTTPT CAGCTCTCT GCGCTCAAGA		
21	ATGGCTCTGA GCTGCTCAGT ATCGAATCG TGAACCTGAC CCGGCCCCAG		
22	ATGCGCTCA GCTCTCAGGT CAGTGGCTTC GCGCTTCAG GAACGGGAT		
23	CAGCGGCTC AGCTGCTGA TACGTTCAA GCGCTTAT? CTGGAGGTCA		
24	GCGCAGCTC TCCCTCTGGG CAACAAGGTA TTTCTGCGT TCGTGGGTG		
25	TTTCCCGGAA ACGTCTCTTT TTTGGGCTAC GCGCTCGGCT TCTTTGTTG		
26	TAGCTCTTPT AATGACTTC TCAAGGTAC GGTGAACGCT CTCATTAATG		
27	CGTTCAGAGA ACATGATATG GGGGTGGGCG TGCTGCGCAC CGGTATTCGC		
28	TGCTTTGCGA TTATGATAG CGAATGATA GGCATGCGG TCGGCAATTT		
29	CTGTGTGAG AAATTCGCG ACAGACTGTA GAGTCTTPT GGGTTTAA		
30	TCAGCGGCA GGGCAATCT GATTTACGCG TAGTACAGC CGTTGCGACG		
31	TTGAGACGTC TCAGCGCTT TCAGAACTC GAGCGTTTFA TGCGCTGCTC		
32	AGCGCGGAT ATGCGCGAG TCCCTGTGCT CAGGTGCGA CTCCTTTTCA		
33	CGGCTACTCT TCCCTCAGC GCGAATATA TCGCATGAG GAGAGGCACT		
34	GCCTTTTCGG CCGGTTTPTA CGCTGAGATG ATAGATGCC ATGCTGTPTT		
35	ATCCGCTGGA AGGGGCGAC TTCTTGAGG AAGTGAAGA AGCTTAGTGG		
36	CGGCTCTGCA AATGAAGAG TTATCAGGGA TTGTAGTGGG ATTTGACCTC		
37	CTCTGCCATC ATGAGCTGTA TCATTCGCTT AGCATTCAGG AGGTAAACAG		
38	CATGAATAAA AGCGAAAAA CAGGAACAT GGCAGCAGA AAGCATGACG		
39	TATTTGTCGG CTTTAAAGTC CGCATGAG CAGCAGAAAC TTATGCTGAG		
40	ACTGACGGAT AAGCAGATA AAACAGCACA GGAATATCAA AGCTGTCTCC		
41	TGCTGATGAA GGTGAAACG GCAGCAGAGA AAGCGCAGA AGCCAGAGCG		
42	AAAGCTAGA ACCTGATGAG GCGGAAAG CAGCGGAG CCGAGGCTGC		
43	CCGTAAAGCC CGTGACCATG CTCTGTACCA GCTGCGCGA TTGCTTATCC		
44	TGCGGGGCTC GGTGACAGT AAGACGGGTA AGCTGTGTA TGATACCGCT		
45	GCTTACTGG GTGCATTAGC CAGTCTGANT GAGCTGAC GGAATATTC		
46	GAGTGGTCA GACTGAAAA TCAGAGGCGA GGAATGCTG AACAGCAAAA		
47	AGTCAGATAG CACCAATATG CAGACCCGCC ATAAAGCCG CTGAGAAAGCC		
48	GTTGACGGCG TTCTCTGTA TTATGGTAG TTCTCTTCA TGATGCCATA		
49	AAAGCGGCTC GTAGTCCAT TTACCCCATC TCATGTCAG AGCCGTGAGC		
50	GAGCGAATC GAATGTCAG AAAAAAGCAG CGACTCAGGT GCTGATGGT		
51	CGAGACAAA AGGAATATTC AGCGATTTGC CGGAGCTTC GAGGCTGCTA		
52	CTTAACTCT TACGCTTTA AGCTCTTCT TGTAGGCGAG CAAAGCTGCT		
53	TTGCGACATC CTTTTTAAT ACTGCGGAAC TGACTAAGT AGTGAATAT		
54	ACACAGGCT GGGATCTATT CTPTTTATCT TTTTATATC TTCTTTATTT		
55	CTATATATTA TAACACTTG AATATAACA AAAAANAC AGAAGGCTCT		
56	AGCGAATTT ACAGAGGCTC TAGCAGAATT TACAGTTT? CCAGCAAAGG		
57	TCAGCAGAA TTACAGATA CCGACAACTC AAGGAAAGG GACTAGTAAT		
58	TATCAITGAG TACCCATCT CAATGGAT? AGGTATAAA ATCACTTGA		
59	CCAATTGAGA TGATATCTG AATTAGTTGT TTTCAAGCA AATGAATAG		
60	CGATTAGCTG CTATGACTTA ACAGAGCATG AAACCAAGCT AATTTTATGC		
61	TGTGTGACG TACTCAACCC CAGCATGAA ACCTTACAA GGTAAAGAGG		
62	GAGCGATGG TTCACTTATA ACCAATACCC TCAGATGATG AACATCAGTA		
63	GGAATAATGC TTATGGTGT TTAGCTAAG CAACAGAGA GCTGATGAG		
64	AGAACTCTGG AATCAGGAA TCCCTTGGTT? AAGGCTTTG AGATTTTCCA		
65	GTGGAACAC TATGCTAGT TCTCAGGTA AATATGAGA TTGTTTCTTA		
66	GTGAAGAGAT ATTGCTTAT CTTTTCCAG? TAAAAAAT? CATAAAAAT		
67	AATCTGGAAC ATGTTAAGTC TTTTGAAGAC AATATCTCTA TGAGGATTTA		
68	TGAGTGTTTA TTAAAGAAC TAACCAAAA GAAACTCTAC AAGGCAATTA		
69	TAGAGATTAG CTTTGTGAA TTTAAGTTCA TGTTAATGCT TGAATAATAC		
70	TACCATGAGT TAAAAAGGCT TAACCAATGG GTTTTGAAGC CAATAGTTAA		
71	AGHTTTAAC ACTTACAGA AATGAATAT GTTGTGTGAT AAGCGAGGCT		
72	GCCCGACTGA TACCTTGATT TTCCAGTTC AACTAGATAG ACAATGTGAT		
73	CTCTTAACCC AACTTGAAGA CAACAGATA AAATGATG GTGACAAAT		
74	ACCAACACCT ATTACATGAG ATTCTTACT AGTAAAGGA CTAGAAAAA		
75	CATCTACGGA TGCTTTAATC GCAAAAATTC AGCTCACAG TTTTGAGGCA		
76	AAATTTTGTG GTGACATGCA AAGTAAAGAT GATCTCAATG GTTCGTTCTC		
77	ATGCTCTGAG CAATAACAC GAACTACACT AGAAGACATA CTGGCTAAT		
78	ACGAGAGAT CTGAGGCTC TATGCTCTT GPACTTATGA GTGAGCATC		
79	AAGACTAACA ACAAAGGTA GAACACTGT TCACCTTATG ATATCAAAGG		
80	GAATACTGTC CATATGACA GATGAAAAAC GTGTAAAAA GATAGATACA		
81	TCAGATCTTT TCGAGTTT? TGCTGCTTT? AAATCTCTC ACCATGACA		
82	GATCGACAAT GTAACAGATG AAGAGCATGT AACACCTAAT AGACAGGTTG		

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or rS
	AAACCGTAA AACAAAGCAA CTAGAACATG AAATGAACA CCGGAGACAA CTTGTTACAG CTCAACAGTC ACACATAGAG AGCCTGAAGC AGCGCATGCT GCTTATCGA TCAAACTGCG GCAACAACGG GAGGCATG AGCCTTCGCG TGGGGAATAA ATCATGGCAA TTCTGGAAGA AATAGCCCTT TCAGCGGCA AACTGAAGC CGGATCTGCG ATCTGATGAA CAACATAGCA ACACGAGAAC AGCCGTTTGG GGGGAGCAA AACCCGTAT TTGGAGTT CGGGGTTT TTTGGGCGA GTGGTGTTCG GGGGTGGCG GCAAGATCCA TTAATGTA CGGGGAGTT TACATCTCAA AACCGCCGCG TTAACACCAT CAGAAATCTT CAGCCGAT TTAGGACCA ACCCTCCGCG GTACACCA AATCATACT GAAAGTCGCT TTGTGAATA AATCGAAT TTGCTGAGTT GAAAGATCAG ATCAAGCATC CTCCCGACAA CACAGACCAT TCCTGCGCAA AGCAAAAGTT CAGATCACC AACTGGTCCA CCTACACAA ACCTCTATC AGCGTGTCT CCCTACCTT CTGCTGGG GGGGAGCAA TTGAGGCTG GTATGAGTGG GCAACACCTT CATACGAGG AAGGCCGAG CGCTATTCTG ATCTGGCAT CACACCGTT CTGGTGATTA AACCGTATT CGGGTGACC CTGGGGCTG CGCGGTTT TATGATTC ATTTTGGCC GATGAGCTT CGGTTGCG TGCCGGATT ACACCACTGT CAGTAAAGCG GCAAGTCGG TTAATGTGAG TTTCAAAACG TCCACCCGCG GTGAAATCG ACACCTGGT ATTGATCCA CGCGCTGA SGTCTTGGT GAAAGCAAT GAAAGTCAG AAACACCGC AAAGAGCGCC GTCCATCTG GCGAAAGTGT CATCTGCTG TTGACAGCAA CACACATGAA GTTGTCTGT CAGACCTGTC GCTGAATAC CTCACGGACT CAGAGCTT CCCGCGCTG ATCCGCGAA CTACAGAAA AATCAAGGCA GCGCGCGAG ACCGGGCTTA CGATACCGCG CTCTGTACG ATGAACCTGG CCGCAAAATA ATCAGCGCG TTATTCTCTC CGGAAAGGT GCGGTTACT GACCTGGTGA ATATGCAAG CTAAAGCTG CAGTGCTTA TCAGGAATG ACCGGATTA ATCCCGTGT GAAATGACA ACAGATTACA ACCGTGCTC GATAGCGGAA ACCGCGATGT ACCGGGTAAA ACAGCTGTT CCGGGTTTAC TGACGCTGCG TACTACGAT GGTCAAGTTG CAGAGGCTAT GCGCTGTGA CGCGCTGA ACAAAATGCG GAAAGAGGT ATCTGTGAA GCGTCTGAT TGCTGAAATA CACAACCGCG TACGGGGAG ACTTACCGA AATCTGATT ATTCAACAAA GCGGCGTGT GTGAATACA AAGCAACCC GTTGAAGTTA TCAGTTGAT GCAATATGCG CAGCGCTAA ATATGCAACA AGACAGGAG CACCTCTGCG ATTAAGCTGT GGTGGTAA CAATATGCTG GGTACATC AGCAGCATG ATGCGCTCCC ACAGCATTCG CTTTGGAGT ATGGAATTC CTGCTGCG GTGCGCGCG TATCAACTCT GAGAGGTG ACCTTGGG GAAATGTGCG CGGAACCCCT ATTTGTTAT TTCTCTAAT ACAATTCAAT ATGATCCCG TCATGAGACA ATACCCCTGA TAAATGCTTC AATATATTG AAAGAGGAG ATATATGTA TCAACATTT CCGTGTGCG CTATTCCTT TTTTTGGCG ATTTCCTCT CCGTTTTTG CTCACCGAGA AACGCTGCTG AAATAAAG ATCTGAAGA TCAGTTGGGT CACGAGTGG GTTACATCGA ACTGATCTC AACAGCGTA AGATCTTGA GAGTTTGG CCGAGAGAC GTTTTCAAT GATGAGACT CTAAAGTTC TGTATGTTGG CGCGGATTA TCGCTGTG ACGCCGGCG AGAGCACTC GGTGCGCGA TACATATTC TCGAATGAC TTGTTGAGT ACTCAAGT CACAGAAAG CATCTACCG ATGCAAGAC AGTAAGGAA TTATGCGTG CTGCAATAC CAGAGTGA AACACTGCG CCACTTACT TCTGACAAG ATCGGAGGAC CGAAGGAGCT AACCGCTTT TTGCACACA TGGGGGATCA TGTAACTGCG CTGATGTT GGGACCGA GTGATATGA GCAATACGA AGCAGAGCT TGAACCGC ATGCTGCG CAATGCGAAC AACGTTGCG CAATATTAA CTGGCGAAT ACTTACTCTA GCTTCCGCG AACAAATTAT AGACTGGATG GAGGCGATA AAGTTGAG ACACCTCTG GGTGCGCG TTGCGGTTG CTGTTTAT GCTATTAAT CTGGAGCGCG TCGAGTGTG TCTCGGCTA CATTTGAGC ACTGGGCA GATGTTAAG CTTCCGTTAT CCGAGTTAC TACAGACCG GGATCAGCG AACTATGAT GAACGAATA GACAGTCC TGAATATG GCTTCACTGA TTACATGCG GTACCGCG ACCAATTTA CTCTATATA CGGACAGCG TCGGACTGT TGTAACTCAG AATAGAAAT GAGGCGCTC ATGCGTTCT GTTCCCTGCT TCACTGGGA AAGGAAAC AACCTGCGC CGCTCTTT GAGCAACGA TCAAAATTA GTGCTGCTC ATCAAAATA TATTCTCAAG ATAAAAACG TTGTGAATA CTGTAAACG T		
68	ATGCGCACGC TTCTGATCGA CAACTACGAC TCGTTACCC AGAAGCTGTT CGATACATC GCGAGGCGCA CCGGGACCC CCGGCTGCTG CCAAGAGAG CGCATGCT GCGGCTGCTC CTGAGGACT TCGAGCGAT CCGTGGCTCC CGCGGCCCG GCAAGCCGGA CGGGGAACG GACTTCGGA TCAGCGCGCG GGGATACCC GACAGCGGCG TCGCGCTCTT GCGCTCTGCT CTGGCCACC AGGATATCG CAGCTCTGCG GCGGAGCCCA TCGACGCGCG GGTCTCGAG GTGCGGACA CCGGAGGA GGTCTCGCG GCGCTCCCT GCGGTTTAC GCGCTGCG TACCACTCC CCGCGGCGAC GAGCTCCCG GAGGAGCTCG AACCCCTGCG CTGGAGCGAC GAGCGGCTG TCATGAGGCT GCGGACCGC GAGAGCGCG TGAAGGAGT CAGTTTCCCA CCGGATCCA TCGGACGGA CTTGGCGCG GAGATCATG CCAACTTCC GAGCTCTGCT CTGCGGACG ACCGGACG TCGGAGCG GCGGCTGCG GCGGAGGCT GCGGAGGCT GCGCTGAGC TGTCTCGGGA CCGGAGAG GAGGCGCG CTGCGTGGC GGCGAGGCG GCGAGTTCT GGTGAGAG CAGCTCCGCT CTGAGAGGCG CTCGCGTT CTCTCTCTC GCGGAGAGC GCGGCGGCT GCGGAGGAT CTCACTACC GCGTGGCA GCGGCGCTC TCGCTGCG GCTCGAGAG CAGCAGAC GCGGAGCGCG GCGGCTCTT CAGCTACCTG GAGGAGCGC TCGAACCGCG GCGGCTGCTC GTCGCGCGC ACCTGCTCT GAGTTTCAAC CTGCGTACG TCGCTACTT CCGCTACGAG CTGAAGCGCG AGACACCGG GAGCCCGCA GTACCGGCG CCGACCGCG CCGCGGCT CTCTCGCGC	three genes (papABC)	Plasmid

WO 02/085923

PCT/US02/12465

SEQ ID #	Sequence	Notes	tRNA or RS
	ACCGCGGCAT GGCCTCGAC CACCAGGAAG GCTGCTGCTA CTTGCTGGCC CTGACCGCC GGGGCCACGA CGACGGGCC CCGCCTTGGC TGGGGGAGAC GGGCGAGACC CTACACGGCC TGGCGTTCG CTTCCCGCCG AGGCGGAGCC CGCCATGGT GTTCGGGTC CCGAGGCGG CGGCGGGTT CCGCCCTGT GCTCGGCAC GCCACGACAA GGACGCTCG GCGCTCCGA ACGGCGAGTC GTACGAGATC TGCCTGACCA ACATGGTCAC CGCGCGACC GAGGCGACGG CCGTGGCGCT CTACTCCGCG CTGGCGCGCA TCGGCCCGT CCGGTCTGGC GCGCTGCTCG AGTTCCTCGA GCTGTCTGGT CTACGCGCT CCGCGAGCG GTTCTTACG ATCGGCGCG AGGCGCGCT CGAGTCCAG CCATCAAGG GGACCGGCC CCGGGGCGA CGGCGGAGG AGGACGAGG GCTCCGCGCC GACCTGGCG GCGGGGAGAA GGAACCGGCC GAGAACCTGA TGATCTCGA CTTGGTCCG AACGACCTCA ACAGGCTCTG CGCGATCGC TCTGTCCACG TCGCCCGGT GTTCGAGTG GAGACCTCG CCGCGTGA CGAGTGTG TCGACCATCC GGGGACGGCT GCGGCGGCG ACCAGACCG CCGCTGCGT ACGCCCGCC TTCCCGGCG GCTCATGAC CGGCGCGCC AGAAGCGAC CCATGGAGAT CATCGACCG CTGAGGAGG GCGCCCGGG CTTCTTACCC GGGGCGCTCG GATGGTTGG CTTCAGCGG GCGCCGAGC CTAGCATGT CATCGGACC ATCTGCTGG CTGACGGCG GCGCGAGTTC GCGTCCGCG GGGCGATGCT GTCCCTCTCC GACCGAGAG AGGAGTTGAG CGAGACCGTG GTCAAGGCC GCGCATGGT CACCGGCTC GACGGGAGCG CAGTGGCGG CGCCGATGA GCGGCTTCCC CCGGAGGCTC GTCTCGCGG GAGCGGAGC GATGGGCGC AGTGTTCGCC GCTGCTTGG GAGGCGGAG AGCCGACGC TCGTGTGGA CTTCGTACCG CCGCCGCGC GCGCGAGCG CTGCTGTGT GCGGACGTC CCGCGCGGG GCGCGAGCT GCGGCGCGCC TCGGGAGCG GGACCTGCT CTGCTCGCC TACAGAGGA CTTGGCGCTC AAGGCGGTG CCGCGGTGAC CCGGCTATG CGACCGGCG CCGTGTGCG CAGACCTGT TCGTCCGA CCGGATGG CCGGAGGTC GCGGCGCAG CCGTGGGCT CGAGACGCT GCGCTCAAC CGAGGTTCC CCGCGCGCG GCGATGACG GCGGCGCGT GCGCGCGTG GTACCGAGG AGGCGCGCG CTTACGCGC CTGCTGCGC TCGTCGAGG GCGCGCGCG AGGCGCGTAC GCGTACGCG GAGGAGACAC GACCGGAGCA GCGCGCGAG CCGTGTGCG AGCGCGCG TGAATCTCT CTTCGCGTC GCGCTGCGC GCGTCCGCT CGAGTTCGG GCGCTGCGG CGACGCGAC GCGCGCGAC CAGGTGCTC TCGCGCTCT GCGCGTGTG CTGCGCGCA GCGCGAGGT GTACGCGGAC ATCGAGCGT CGAACCTCG GCGCGCTCC GCGCGCGCG CCGTGTGCG GCGCGCGCG TCTTTCGCG CCGTGTGCG CAGGACCG GCGCGCGCG AGGAGTCCGA CCGCGCGAC GACCGGAGC GACCGGAGC CCGCGCGCAT CCGCGCGGAT GCGCGCGCG CCGGAGCTC GACCGGCTT TCGAGGAACT CCGCGCGCTC ATGGAGCGG AGCTCGCGC GCGCGAGGAC CATGCGGAG AGCTGTTCG CACCTTCAC CCGACGAGC AGGAGGCGA GAGGCGCGA TGACCGAGCA GAGCGAGCT CAGGTTGCG CTGCGCGCG AGCTGAGCG CCGCGAGCG AGCGTCTCG ACACGCGCG GCGCGCGAT GACTGAGCG GCGCGGTCG GCGGTACAG TCGCGGAGC GCGTCCGAT GATGAGCGC GCGCGGTCG GCGTGTCAA GAGACGCGC GCGCGTACG CCGCGAGCA CCGCGTTCG GATGCTTCC TGATGAAGT CTACGAGTG ATCAAGCG AGATGTTCG CCGCGAGGAC CTGCTGATGA GCGCGGAGG CCGCGCGCG GAGGACGCGC GGTGA		

WO 02/085923

PCT/US02/12465

[385] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described
5 above can be used in various combinations. All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

WO 02/085923

PCT/US02/12465

WHAT IS CLAIMED IS:

1. A composition comprising a translation system, the translation system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl tRNA synthetase (O-RS), wherein the O-RS preferentially aminoacylates the O-tRNA with at least one
5 unnatural amino acid in the translation system and the O-tRNA recognizes at least one selector codon.
2. The composition of claim 1, wherein the translation system comprises a cell.
3. The composition of claim 2, wherein the translation system comprises a bacterial cell.
- 10 4. The composition of claim 3, wherein the translation system comprises an Escherichia coli cell.
5. The composition of claim 2, wherein the translation system comprises an archaeobacterial cell.
- 15 6. The composition of claim 2, wherein the translation system comprises a eukaryotic cell.
7. The composition of claim 6, wherein the eukaryotic cell comprises a yeast cell, a mammalian cell, a plant cell, or an insect cell.
8. The composition of claim 1, wherein the translation system comprises an in vitro translation system.
- 20 9. The composition of claim 8, wherein the translation system comprises a cellular extract.
10. The composition of claim 1, wherein the O-tRNA comprises a nucleic acid comprising a polynucleotide sequence selected from the group consisting of: SEQ ID NO:1-3 and a complementary polynucleotide sequence thereof.
- 25 11. The composition of claim 1, wherein the O-RS comprises a polypeptide selected from the group consisting of:
a polypeptide comprising an amino acid sequence selected from the group consisting of
SEQ ID NO: 4-34

WO 02/085923

PCT/US02/12465

a polypeptide encoded by a nucleic acid comprising a polynucleotide sequence selected from the group consisting of: SEQ ID NO:35-66 and a complementary polynucleotide sequence thereof.

12. The composition of claim 1, wherein the at least one unnatural amino acid is selected from the group consisting of: an O-methyl-L-tyrosine, an L-3-(2-naphthyl)alanine, a 3-methyl-phenylalanine, an O-4-allyl-L-tyrosine, a 4-propyl-L-tyrosine, a tri-O-acetyl-GlcNAc β -serine, an L-Dopa, a fluorinated phenylalanine, an isopropyl-L-phenylalanine, a *p*-azido-L-phenylalanine, a *p*-acyl-L-phenylalanine, a *p*-benzoyl-L-phenylalanine, an L-phosphoserine, a phosphoserine, a phosphotyrosine, a *p*-iodo-phenylalanine, a *p*-bromophenylalanine, a *p*-amino-L-phenylalanine, an isopropyl-L-phenylalanine, an unnatural analogue of a tyrosine amino acid; an unnatural analogue of a glutamine amino acid; an unnatural analogue of a phenylalanine amino acid; an unnatural analogue of a serine amino acid; an unnatural analogue of a threonine amino acid; an alkyl, aryl, acyl, azido, cyano, halo, hydrazine, hydrazide, hydroxyl, alkenyl, alkynyl, ether, thiol, sulfonyl, seleno, ester, thioacid, borate, boronate, phospho, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, hydroxylamine, keto, or amino substituted amino acid, or any combination thereof; an amino acid with a photoactivatable cross-linker; a spin-labeled amino acid; a fluorescent amino acid; an amino acid with a novel functional group; an amino acid that covalently or noncovalently interacts with another molecule; a metal binding amino acid; a metal-containing amino acid; a radioactive amino acid; a photocaged and/or photoisomerizable amino acid; a biotin or biotin-analogue containing amino acid; a glycosylated or carbohydrate modified amino acid; a keto containing amino acid; amino acids comprising polyethylene glycol or polyether; a heavy atom substituted amino acid; a chemically cleavable or photocleavable amino acid; an amino acid with an elongated side chain; an amino acid containing a toxic group; a sugar substituted amino acid, e.g., a sugar substituted serine or the like; a carbon-linked sugar-containing amino acid; a redox-active amino acid; an α -hydroxy containing acid; an amino thio acid containing amino acid; an α,α disubstituted amino acid; a β -amino acid; and a cyclic amino acid other than proline.
13. The composition of claim 12, wherein the at least one unnatural amino acid is an O-methyl-L-tyrosine.

WO 02/085923

PCT/US02/12465

14. The composition of claim 12, wherein the at least one unnatural amino acid is an L-3-(2-naphthyl)alanine.
15. The composition of claim 12, wherein the at least one unnatural amino acid is an amino-, isopropyl-, or O-allyl-containing phenylalanine analogue.
- 5 16. The composition of claim 1, wherein the at least one selector codon is a nonsense codon, a rare codon, or a four base codon.
17. The composition of claim 16, wherein the at least one selector codon is an amber codon.
18. An Escherichia coli cell comprising a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ and a mutant TyrRS (LWJ16), wherein the mutant TyrRS (LWJ16) preferentially aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with O-methyl-L-tyrosine in the cell and the cell uses the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ to recognize an amber codon.
- 10 19. An Escherichia coli cell comprising a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ and an SS12-TyrRS, wherein the SS12-TyrRS preferentially aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with L-3-(2-naphthyl)alanine in the cell and the cell uses the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ to recognize an amber codon.
- 15 20. A composition comprising a protein, wherein the protein comprises at least one unnatural amino acid, and the protein is present in a concentration of at least about 100 micrograms per liter.
21. The composition of claim 20, wherein the protein is homologous to a
- 20 therapeutic protein selected from the group consisting of a cytokine, a growth factor, a growth factor receptor, an interferon, an interleukin, an inflammatory molecule, an oncogene product, a peptide hormone, a signal transduction molecule, a steroid hormone receptor, a transcriptional activator, a transcriptional suppressor, erythropoietin (EPO), insulin, human growth hormone, epithelial Neutrophil Activating Peptide-78,
- 25 $\text{GRO}\alpha/\text{MGSA}$, $\text{GRO}\beta$, $\text{GRO}\gamma$, $\text{MIP-1}\alpha$, $\text{MIP-1}\delta$, MCP-1 , hepatocyte growth factor, insulin-like growth factor, leukemia inhibitory factor, oncostatin M, PD-ECSF, PDGF, pleiotropin, SCF, c-kit ligand, VEGF, G-CSF, IL-1, IL-2, IL-8, IGF-I, IGF-II, FGF (fibroblast growth factor), PDGF, TNF, $\text{TGF-}\alpha$, $\text{TGF-}\beta$, EGF (epidermal growth factor), KGF (keratinocyte growth factor), SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-
- 30 I/LFA-1, hyalurin/CD44, Mos, Ras, Raf, Met; p53, Tat, Fos, Myc, Jun, Myb, Rel.

WO 02/085923

PCT/US02/12465

estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

22. The composition of claim 20, wherein the protein is homologous to a therapeutic protein selected from the group consisting of: an Alpha-1 antitrypsin, an
- 5 Angiostatin, an Antihemolytic factor, an antibody, an Apolipoprotein, an Apoprotein, an Atrial natriuretic factor, an Atrial natriuretic polypeptide, an Atrial peptide, a C-X-C chemokine, T39765, NAP-2, ENA-78, a Gro-a, a Gro-b, a Gro-c, an IP-10, a GCP-2, an NAP-4, an SDF-1, a PF4, a MIG, a Calcitonin, a c-kit ligand, a cytokine, a CC chemokine, a Monocyte chemoattractant protein-1, a Monocyte chemoattractant protein-2, a Monocyte
- 10 chemoattractant protein-3, a Monocyte inflammatory protein-1 alpha, a Monocyte inflammatory protein-1 beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262, a CD40, a CD40 ligand, a C-kit Ligand, a Collagen, a Colony stimulating factor (CSF), a Complement factor 5a, a Complement inhibitor, a Complement receptor 1, a cytokine, an epithelial Neutrophil Activating Peptide-78, a GRO α /MGSA, a GRO β , a
- 15 GRO γ , a MIP-1 α , a MIP-1 β , a MCP-1, an Epidermal Growth Factor (EGF), an epithelial Neutrophil Activating Peptide, an Erythropoietin (EPO), an Exfoliating toxin, a Factor IX, a Factor VII, a Factor VIII, a Factor X, a Fibroblast Growth Factor (FGF), a Fibrinogen, a Fibronectin, a G-CSF, a GM-CSF, a Glucocerebrosidase, a Gonadotropin, a growth factor, a growth factor receptor, a Hedgehog protein, a Hemoglobin, a Hepatocyte Growth Factor
- 20 (HGF), a Hirudin, a Human serum albumin, an ICAM-1, an ICAM-1 receptor, an LFA-1, an LFA-1 receptor, an Insulin, an Insulin-like Growth Factor (IGF), an IGF-I, an IGF-II, an interferon, an IFN- α , an IFN- β , an IFN- γ , an interleukin, an IL-1, an IL-2, an IL-3, an IL-4, an IL-5, an IL-6, an IL-7, an IL-8, an IL-9, an IL-10, an IL-11, an IL-12, a Keratinocyte Growth Factor (KGF), a Lactoferrin, a leukemia inhibitory factor, a
- 25 Luciferase, a Neurturin, a Neutrophil inhibitory factor (NIF), an oncostatin M, an Osteogenic protein, an oncogene product, a Parathyroid hormone, a PD-ECSF, a PDGF, a peptide hormone, a Human Growth Hormone, a Pleiotropin, a Protein A, a Protein G, a Pyrogenic exotoxins A, B, or C, a Relaxin, a Renin, an SCF, a Soluble complement receptor I, a Soluble I-CAM 1, a Soluble interleukin receptors, a Soluble TNF receptor, a
- 30 Somatomedin, a Somatostatin, a Somatotropin, a Streptokinase, a Superantigens, a Staphylococcal enterotoxins, an SEA, an SEB, an SEC1, an SEC2, an SEC3, an SED, an SEE, a steroid hormone receptor, a Superoxide dismutase, a Toxic shock syndrome toxin, a Thymosin alpha 1, a Tissue plasminogen activator, a tumor growth factor (TGF), a TGF-

WO 02/085923

PCT/US02/12465

- α , a TGF- β , a Tumor Necrosis Factor, a Tumor Necrosis Factor alpha, a Tumor necrosis factor beta, a Tumor necrosis factor receptor (TNFR), a VLA-4 protein, a VCAM-1 protein, a Vascular Endothelial Growth Factor (VEGEF), a Urokinase, a Mos, a Ras, a Raf, a Met; a p53, a Tat, a Fos, a Myc, a Jun, a Myb, a Rel, an estrogen receptor, a progesterone receptor, a testosterone receptor, an aldosterone receptor, an LDL receptor, and a corticosterone.
23. The composition of claim 22, wherein the protein is at least about 50% or at least about 75% identical to the therapeutic protein.
24. The composition of claim 22, wherein the therapeutic protein is erythropoietin (EPO).
25. The composition of claim 20, wherein the protein comprises at least two unnatural amino acids.
26. The composition of claim 20, wherein the protein comprises at least three unnatural amino acids.
27. The composition of claim 20, wherein the protein comprises at least four unnatural amino acids.
28. The composition of claim 20, wherein the protein comprises at least five or more unnatural amino acids.
29. The composition of claim 20, wherein the composition comprises at least about 100 micrograms of protein.
30. The composition of claim 20, wherein the protein is DHFR, and the at least one unnatural amino acid is selected from the group consisting of O-methyl-L-tyrosine and L-3-(2-naphthyl)alanine.
31. A composition comprising a protein, wherein the protein comprises at least one unnatural amino acid, and the protein is homologous to a therapeutic protein selected from the group consisting of a cytokine, a growth factor, a growth factor receptor, an interferon, an interleukin, an inflammatory molecule, an oncogene product, a peptide hormone, a signal transduction molecule, a steroid hormone receptor, a transcriptional activator, a transcriptional suppressor, erythropoietin (EPO), insulin, human growth hormone, epithelial Neutrophil Activating Peptide-78, GRO α /MGSA, GRO β , GRO γ , MIP-1 α , MIP-1 δ , MCP-1, hepatocyte growth factor, insulin-like growth factor, leukemia

WO 02/085923

PCT/US02/12465

- inhibitory factor, oncostatin M, PD-ECSF, PDGF, pleiotropin, SCF, c-kit ligand, VEGF, G-CSF, IL-1, IL-2, IL-8, IGF-I, IGF-II, FGF (fibroblast growth factor), PDGF, TNF, TGF- α , TGF- β , EGF (epidermal growth factor), KGF (keratinocyte growth factor), SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, hyalurin/CD44, Mos, Ras, Raf,
- 5 Met; p53, Tat, Fos, Myc, Jun, Myb, Rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.
32. A composition comprising a protein, wherein the protein comprises at least one unnatural amino acid, and the protein is homologous to a therapeutic protein selected from the group consisting of a an Alpha-1 antitrypsin, an Angiostatin, an Antihemolytic
- 10 factor, an antibody, an Apolipoprotein, an Apoprotein, an Atrial natriuretic factor, an Atrial natriuretic polypeptide, an Atrial peptide, a C-X-C chemokine, T39765, NAP-2, ENA-78, a Gro-a, a Gro-b, a Gro-c, an IP-10, a GCP-2, an NAP-4, an SDF-1, a PF4, a MIG, a Calcitonin, a c-kit ligand, a cytokine, a CC chemokine, a Monocyte chemoattractant protein-1, a Monocyte chemoattractant protein-2, a Monocyte
- 15 chemoattractant protein-3, a Monocyte inflammatory protein-1 alpha, a Monocyte inflammatory protein-1 beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262, a CD40, a CD40 ligand, a C-kit Ligand, a Collagen, a Colony stimulating factor (CSF), a Complement factor 5a, a Complement inhibitor, a Complement receptor 1, a cytokine, an epithelial Neutrophil Activating Peptide-78, a GRO α /MGSA, a GRO β , a
- 20 GRO γ , a MIP-1 α , a MIP-1 β , a MCP-1, an Epidermal Growth Factor (EGF), an epithelial Neutrophil Activating Peptide, an Erythropoietin (EPO), an Exfoliating toxin, a Factor IX, a Factor VII, a Factor VIII, a Factor X, a Fibroblast Growth Factor (FGF), a Fibrinogen, a Fibronectin, a G-CSF, a GM-CSF, a Glucocerebrosidase, a Gonadotropin, a growth factor, a growth factor receptor, a Hedgehog protein, a Hemoglobin, a Hepatocyte Growth Factor
- 25 (HGF), a Hirudin, a Human serum albumin, an ICAM-1, an ICAM-1 receptor, an LFA-1, an LFA-1 receptor, an Insulin, an Insulin-like Growth Factor (IGF), an IGF-I, an IGF-II, an interferon, an IFN- α , an IFN- β , an IFN- γ , an interleukin, an IL-1, an IL-2, an IL-3, an IL-4, an IL-5, an IL-6, an IL-7, an IL-8, an IL-9, an IL-10, an IL-11, an IL-12, a Keratinocyte Growth Factor (KGF), a Lactoferrin, a leukemia inhibitory factor, a
- 30 Luciferase, a Neurturin, a Neutrophil inhibitory factor (NIF), an oncostatin M, an Osteogenic protein, an oncogene product, a Parathyroid hormone, a PD-ECSF, a PDGF, a peptide hormone, a Human Growth Hormone, a Pleiotropin, a Protein A, a Protein G, a Pyrogenic exotoxins A, B, or C, a Relaxin, a Renin, an SCF, a Soluble complement

WO 02/085923

PCT/US02/12465

receptor I, a Soluble I-CAM 1, a Soluble interleukin receptors, a Soluble TNF receptor, a Somatomedin, a Somatostatin, a Somatotropin, a Streptokinase, a Superantigens, a Staphylococcal enterotoxins, an SEA, an SEB, an SEC1, an SEC2, an SEC3, an SED, an SEE, a steroid hormone receptor, a Superoxide dismutase, a Toxic shock syndrome toxin, a Thymosin alpha 1, a Tissue plasminogen activator, a tumor growth factor (TGF), a TGF- α , a TGF- β , a Tumor Necrosis Factor, a Tumor Necrosis Factor alpha, a Tumor necrosis factor beta, a Tumor necrosis factor receptor (TNFR), a VLA-4 protein, a VCAM-1 protein, a Vascular Endothelial Growth Factor (VEGFR), a Urokinase, a Mos, a Ras, a Raf, a Met, a p53, a Tat, a Fos, a Myc, a Jun, a Myb, a Rel, an estrogen receptor, a progesterone receptor, a testosterone receptor, an aldosterone receptor, an LDL receptor, and a corticosterone.

33. The composition of claim 32, wherein the protein comprises at least two unnatural amino acids.

34. The composition of claim 32, wherein the protein comprises at least three unnatural amino acids.

35. The composition of claim 32, wherein the protein comprises at least four unnatural amino acids.

36. The composition of claim 32, wherein the protein comprises at least five or more unnatural amino acids.

37. The composition of claim 32, wherein the protein is at least about 50% or at least about 75% identical to the therapeutic protein.

38. The composition of claim 32, wherein the therapeutic protein is erythropoietin (EPO).

39. A composition comprising a protein and a pharmaceutically acceptable excipient, wherein the protein comprises at least one unnatural amino acid.

40. The composition of claim 39, wherein the protein comprises at least two unnatural amino acids.

41. The composition of claim 39, wherein the protein comprises at least three unnatural amino acids.

42. The composition of claim 39, wherein the protein comprises at least four unnatural amino acids.

WO 02/085923

PCT/US02/12465

43. The composition of claim 39, wherein the protein comprises at least five or more unnatural amino acids.

44. The composition of claim 39, wherein the protein is homologous to a therapeutic protein selected from the group consisting of a cytokine, a growth factor, a growth factor receptor, an interferon, an interleukin, an inflammatory molecule, an oncogene product, a peptide hormone, a signal transduction molecule, a steroid hormone receptor, a transcriptional activator, a transcriptional suppressor, erythropoietin (EPO), insulin, human growth hormone, epithelial Neutrophil Activating Peptide-78, GRO α /MGSA, GRO β , GRO γ , MIP-1 α , MIP-1 δ , MCP-1, hepatocyte growth factor, 5 insulin-like growth factor, leukemia inhibitory factor, oncostatin M, PD-ECSF, PDGF, pleiotropin, SCF, c-kit ligand, VEGF, G-CSF, IL-1, IL-2, IL-8, IGF-I, IGF-II, FGF (fibroblast growth factor), PDGF, TNF, TGF- α , TGF- β , EGF (epidermal growth factor), KGF (keratinocyte growth factor), SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, hyaluron/CD44, Mos, Ras, Raf, Met, p53, Tat, Fos, Myc, Jun, Myb, Rel, 10 estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

45. The composition of claim 39, wherein the protein is homologous to a therapeutic protein selected from the group consisting of an Alpha-1 antitrypsin, an Angiostatin, an Antihemolytic factor, an antibody, an Apolipoprotein, an Apoprotein, an 20 Atrial natriuretic factor, an Atrial natriuretic polypeptide, an Atrial peptide, a C-X-C chemokine, T39765, NAP-2, ENA-78, a Gro-a, a Gro-b, a Gro-c, an IP-10, a GCP-2, an NAP-4, an SDF-1, a PF4, a MIG, a Calcitonin, a c-kit ligand, a cytokine, a CC chemokine, a Monocyte chemoattractant protein-1, a Monocyte chemoattractant protein-2, a Monocyte chemoattractant protein-3, a Monocyte inflammatory protein-1 alpha, a Monocyte 25 inflammatory protein-1 beta, RANTES, I309, R83915, R91733, HCC1, TS8847, D31065, T64262, a CD40, a CD40 ligand, a C-kit Ligand, a Collagen, a Colony stimulating factor (CSF), a Complement factor 5a, a Complement inhibitor, a Complement receptor 1, a cytokine, an epithelial Neutrophil Activating Peptide-78, a GRO α /MGSA, a GRO β , a GRO γ , a MIP-1 α , a MIP-1 δ , a MCP-1, an Epidermal Growth Factor (EGF), an epithelial 30 Neutrophil Activating Peptide, an Erythropoietin (EPO), an Exfoliating toxin, a Factor IX, a Factor VII, a Factor VIII, a Factor X, a Fibroblast Growth Factor (FGF), a Fibrinogen, a Fibronectin, a G-CSF, a GM-CSF, a Glucocerebrosidase, a Gonadotropin, a growth factor, a growth factor receptor, a Hedgehog protein, a Hemoglobin, a Hepatocyte Growth Factor

WO 02/085923

PCT/US02/12465

- (HGF), a Hirudin, a Human serum albumin, an ICAM-1, an ICAM-1 receptor, an LFA-1, an LFA-1 receptor, an Insulin, an Insulin-like Growth Factor (IGF), an IGF-I, an IGF-II, an interferon, an IFN- α , an IFN- β , an IFN- γ , an interleukin, an IL-1, an IL-2, an IL-3, an IL-4, an IL-5, an IL-6, an IL-7, an IL-8, an IL-9, an IL-10, an IL-11, an IL-12, a
- 5 Keratinocyte Growth Factor (KGF), a Lactoferrin, a leukemia inhibitory factor, a Luciferase, a Neurturin, a Neutrophil inhibitory factor (NIF), an oncostatin M, an Osteogenic protein, an oncogene product, a Parathyroid hormone, a PD-ECSF, a PDGF, a peptide hormone, a Human Growth Hormone, a Pleiotropin, a Protein A, a Protein G, a Pyrogenic exotoxins A, B, or C, a Relaxin, a Renin, an SCF, a Soluble complement
- 10 receptor I, a Soluble I-CAM 1, a Soluble interleukin receptors, a Soluble TNF receptor, a Somatomedin, a Somatostatin, a Somatotropin, a Streptokinase, a Superantigens, a Staphylococcal enterotoxins, an SEA, an SEB, an SEC1, an SEC2, an SEC3, an SED, an SEE, a steroid hormone receptor, a Superoxide dismutase, a Toxic shock syndrome toxin, a Thymosin alpha 1, a Tissue plasminogen activator, a tumor growth factor (TGF), a TGF-
- 15 α , a TGF- β , a Tumor Necrosis Factor, a Tumor Necrosis Factor alpha, a Tumor necrosis factor beta, a Tumor necrosis factor receptor (TNFR), a VLA-4 protein, a VCAM-1 protein, a Vascular Endothelial Growth Factor (VEGEGF), a Urokinase, a Mos, a Ras, a Raf, a Met, a p53, a Tat, a Fos, a Myc, a Jun, a Myb, a Rel, an estrogen receptor, a progesterone receptor, a testosterone receptor, an aldosterone receptor, an LDL receptor,
- 20 and a corticosterone.

46. The composition of claim 45, wherein the protein is at least about 50% or at least about 75% identical to the therapeutic protein.

47. The composition of claim 45, wherein the therapeutic protein is erythropoietin (EPO).

- 25 48. A method for producing in a translation system at least one protein comprising at least one unnatural amino acid, the method comprising:
- providing the translation system with at least one nucleic acid comprising at least one selector codon, wherein the nucleic acid encodes the at least one protein;
- providing the translation system with an orthogonal tRNA (O-tRNA), wherein the O-
- 30 tRNA functions in the translation system and wherein the O-tRNA recognizes the at least one selector codon;

WO 02/085923

PCT/US02/12465

providing the translation system with an orthogonal aminoacyl tRNA synthetase (O-RS),
 wherein the O-RS preferentially aminoacylates the O-tRNA with the at least one unnatural
 amino acid in the translation system; and

providing the translation system with the at least one unnatural amino, thereby producing
 5 in the translation system the at least one protein comprising the at least one unnatural
 amino acid.

49. The method of claim 48, wherein the translation system comprises a cell.

50. The method of claim 49, wherein the translation system comprises a bacterial
 cell.

10 51. The method of claim 50, wherein the translation system comprises an
 Escherichia coli cell.

52. The method of claim 49, wherein the translation system comprises an
 archaeobacterial cell.

15 53. The method of claim 49, wherein the translation system comprises a
 eukaryotic cell.

54. The method of claim 53, wherein the eukaryotic cell comprises a yeast cell, a
 mammalian cell, a plant cell, or an insect cell.

20 55. The protein comprising at least one unnatural amino acid produced by the
 method of claim 49, wherein the protein is processed and modified in a cell-dependent
 manner.

56. The protein of claim 55, wherein the protein is homologous to a therapeutic
 protein selected from the group consisting of a cytokine, a growth factor, a growth factor
 receptor, an interferon, an interleukin, an inflammatory molecule, an oncogene product, a
 peptide hormone, a signal transduction molecule, a steroid hormone receptor, a
 25 transcriptional activator, a transcriptional suppressor, erythropoietin (EPO), insulin,
 human growth hormone, epithelial Neutrophil Activating Peptide-78, GRO α /MGSA,
 GRO β , GRO γ , MIP-1 α , MIP-1 β , MCP-1, hepatocyte growth factor, insulin-like growth
 factor, leukemia inhibitory factor, oncostatin M, PD-ECSF, PDGF, pleiotropin, SCF, c-kit
 ligand, VEGF, G-CSF, IL-1, IL-2, IL-8, IGF-I, IGF-II, FGF (fibroblast growth factor),

WO 02/085923

PCT/US02/12465

PDGF, TNF, TGF- α , TGF- β , EGF (epidermal growth factor), KGF (keratinocyte growth factor), SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, hyalurin/CD44, Mos, Ras, Raf, Met; p53, Tat, Fos, Myc, Jun, Myb, Rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

- 5 57. The protein of claim 55, wherein the protein is homologous to a therapeutic protein selected from the group consisting of a an Alpha-1 antitrypsin, an Angiostatin, an Antihemolytic factor, an antibody, an Apolipoprotein, an Apoprotein, an Atrial natriuretic factor, an Atrial natriuretic polypeptide, an Atrial peptide, a C-X-C chemokine, T39765, NAP-2, ENA-78, a Gro-a, a Gro-b, a Gro-c, an IP-10, a GCP-2, an NAP-4, an SDF-1, a
- 10 PF4, a MIG, a Calcitonin, a c-kit ligand, a cytokine, a CC chemokine, a Monocyte chemoattractant protein-1, a Monocyte chemoattractant protein-2, a Monocyte chemoattractant protein-3, a Monocyte inflammatory protein-1 alpha, a Monocyte inflammatory protein-1 beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262, a CD40, a CD40 ligand, a C-kit Ligand, a Collagen, a Colony stimulating factor
- 15 (CSF), a Complement factor 5a, a Complement inhibitor, a Complement receptor 1, a cytokine, an epithelial Neutrophil Activating Peptide-78, a GRO α /MGSA, a GRO β , a GRO γ , a MIP-1 α , a MIP-1 β , a MCP-1, an Epidermal Growth Factor (EGF), an epithelial Neutrophil Activating Peptide, an Erythropoietin (EPO), an Exfoliating toxin, a Factor IX, a Factor VII, a Factor VIII, a Factor X, a Fibroblast Growth Factor (FGF), a Fibrinogen, a
- 20 Fibronectin, a G-CSF, a GM-CSF, a Glucocerebrosidase, a Gonadotropin, a growth factor, a growth factor receptor, a Hedgehog protein, a Hemoglobin, a Hepatocyte Growth Factor (HGF), a Hirudin, a Human serum albumin, an ICAM-1, an ICAM-1 receptor, an LFA-1, an LFA-1 receptor, an Insulin, an Insulin-like Growth Factor (IGF), an IGF-I, an IGF-II, an interferon, an IFN- α , an IFN- β , an IFN- γ , an interleukin, an IL-1, an IL-2, an IL-3, an
- 25 IL-4, an IL-5, an IL-6, an IL-7, an IL-8, an IL-9, an IL-10, an IL-11, an IL-12, a Keratinocyte Growth Factor (KGF), a Lactoferrin, a leukemia inhibitory factor, a Luciferase, a Neurturin, a Neutrophil inhibitory factor (NIF), an oncostatin M, an Osteogenic protein, an oncogene product, a Parathyroid hormone, a PD-ECSF, a PDGF, a peptide hormone, a Human Growth Hormone, a Pleiotropin, a Protein A, a Protein G, a
- 30 Pyrogenic exotoxins A, B, or C, a Relaxin, a Renin, an SCF, a Soluble complement receptor I, a Soluble I-CAM 1, a Soluble interleukin receptors, a Soluble TNF receptor, a Somatomedin, a Somatostatin, a Somatotropin, a Streptokinase, a Superantigens, a Staphylococcal enterotoxins, an SEA, an SEB, an SEC1, an SEC2, an SEC3, an SED, an

WO 02/085923

PCT/US02/12465

SEE, a steroid hormone receptor, a Superoxide dismutase, a Toxic shock syndrome toxin, a Thymosin alpha 1, a Tissue plasminogen activator, a tumor growth factor (TGF), a TGF- α , a TGF- β , a Tumor Necrosis Factor, a Tumor Necrosis Factor alpha, a Tumor necrosis factor beta, a Tumor necrosis factor receptor (TNFR), a VLA-4 protein, a VCAM-1

5 protein, a Vascular Endothelial Growth Factor (VEGEGF), a Urokinase, a Mos, a Ras, a Raf, a Met; a p53, a Tat, a Fos, a Myc, a Jun, a Myb, a Rel, an estrogen receptor, a progesterone receptor, a testosterone receptor, an aldosterone receptor, an LDL receptor, and a corticosterone.

58. The protein of claim 57, wherein the protein is at least about 50% or at least
10 about 75% identical to the therapeutic protein.

59. The protein of claim 57, wherein the therapeutic protein is erythropoietin (EPO).

60. The protein of claim 55, wherein the protein comprises at least two unnatural amino acids.

15 61. The protein of claim 55, wherein the protein comprises at least three unnatural amino acids.

62. The protein of claim 55, wherein the protein comprises at least four unnatural amino acids.

20 63. The protein of claim 55, wherein the protein comprises at least five or more unnatural amino acids.

64. The method of claim 48, wherein the translation system comprises an in vitro translation system.

65. The method of claim 64, wherein the translation system comprises a cell extract.

25 66. The method of claim 48 wherein the at least one protein is a Asp112TAG mutant of chloramphenicol acetyltransferase (CAT), a Tyr163TAG mutant of mouse dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse dihydrofolate reductase (DHFR) comprising a COOH His6tag.

30 67. The method of claim 48 wherein the unnatural amino acid is selected from the group consisting of: an O-methyl-L-tyrosine, an L-3-(2-naphthyl)alanine, a 3-methyl-phenylalanine, an O-4-allyl-L-tyrosine, a 4-propyl-L-tyrosine, a tri-O-acetyl-GlcNAc β -

WO 02/085923

PCT/US02/12465

serine, an L-Dopa, a fluorinated phenylalanine, an isopropyl-L-phenylalanine, a *p*-azido-L-phenylalanine, a *p*-acyl-L-phenylalanine, a *p*-benzoyl-L-phenylalanine, an L-phosphoserine, a phosphoserine, a phosphotyrosine, a *p*-iodo-phenylalanine, a *p*-bromophenylalanine, a *p*-amino-L-phenylalanine, an isopropyl-L-phenylalanine, an

- 5 unnatural analogue of a tyrosine amino acid; an unnatural analogue of a glutamine amino acid; an unnatural analogue of a phenylalanine amino acid; an unnatural analogue of a serine amino acid; an unnatural analogue of a threonine amino acid; an alkyl, aryl, acyl, azido, cyano, halo, hydrazine, hydrazide, hydroxyl, alkenyl, alkynyl, ether, thiol, sulfonyl, seleno, ester, thioacid, borate, boronate, phospho, phosphono, phosphine, heterocyclic,
- 10 enone, imine, aldehyde, hydroxylamine, keto, or amino substituted amino acid, or any combination thereof; an amino acid with a photoactivatable cross-linker; a spin-labeled amino acid; a fluorescent amino acid; an amino acid with a novel functional group; an amino acid that covalently or noncovalently interacts with another molecule; a metal binding amino acid; a metal-containing amino acid; a radioactive amino acid; a
- 15 photocaged and/or photoisomerizable amino acid; a biotin or biotin-analogue containing amino acid; a glycosylated or carbohydrate modified amino acid; a keto containing amino acid; amino acids comprising polyethylene glycol or polyether; a heavy atom substituted amino acid; a chemically cleavable or photocleavable amino acid; an amino acid with an elongated side chain; an amino acid containing a toxic group; a sugar substituted amino
- 20 acid, e.g., a sugar substituted serine or the like; a carbon-linked sugar-containing amino acid; a redox-active amino acid; an α -hydroxy containing acid; an amino thio acid containing amino acid; an α,α disubstituted amino acid; a β -amino acid; and a cyclic amino acid other than proline.

68. The method of claim 48 wherein the at least one unnatural amino acid is a O-methyl-L-tyrosine.

69. The method of claim 48 wherein the at least one unnatural amino acid is an L-3-(2-naphthyl)alanine.

70. The method of claim 48 wherein the at least one unnatural amino acid is an amino-, isopropyl-, or O-allyl-containing phenylalanine analogue.

71. The method of claim 48, wherein the O-tRNA comprises a nucleic acid comprising a polynucleotide sequence selected from the group consisting of: SEQ ID NO:1-3 and a complementary polynucleotide sequence thereof.

WO 02/085923

PCT/US02/12465

72. The method of claim 48 wherein the at least one selector codon is nonsense codon, a rare codon, or a four base codon.

73. The method of claim 72 wherein the nonsense codon is an amber codon.

74. The method of claim 48, wherein the O-RS preferentially aminoacylates the
5 O-tRNA with an O-methyl-L-tyrosine.

75. The method of claim 48, wherein the O-RS preferentially aminoacylates the O-tRNA with an amino-, isopropyl-, or O-allyl-containing phenylalanine analogue.

76. The method of claim 48, wherein the O-RS comprises a polypeptide selected from the group consisting of:
10 a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 4-34; and,
a polypeptide encoded by a nucleic acid comprising a polynucleotide sequence selected from the group consisting of: SEQ ID NO:35-66 and a complementary polynucleotide sequence thereof.

15 77. The method of claim 48 wherein the unnatural amino acid is provided exogenously.

78. The method of claim 48 wherein the translation system is a cell and the unnatural amino acid is biosynthesized by the cell.

79. A method for producing in an Escherichia coli cell at least one protein
20 comprising at least one O-methyl-L-tyrosine, the method comprising:
providing the translation system with at least one nucleic acid comprising an amber codon, wherein the nucleic acid encodes the at least one protein;
providing the translation system with a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$, wherein the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ functions in the cell and wherein the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ recognizes the amber codon;
25 providing the translation system with a mutant TyrRS (LWJ16), wherein the mutant TyrRS (LWJ16) aminoacylates the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with the O-methyl-L-tyrosine in the cell;
and

WO 02/085923

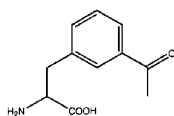
PCT/US02/12465

providing the cell with the O-methyl-L-tyrosine, thereby producing in the cell at least one protein comprising the O-methyl-L-tyrosine.

- 80.** A method for producing in an Escherichia coli cell at least one protein comprising at least one L-3-(2-naphthyl)alanine, the method comprising:
- 5 providing the translation system with at least one nucleic acid comprising an amber codon, wherein the nucleic acid encodes the at least one protein;
- providing the cell with a $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$, wherein the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ functions in the cell and wherein the $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ recognizes the amber codon;
- providing the cell with an SS12-TyrRS, wherein the SS12-TyrRS aminoacylates the
- 10 $\text{mtRNA}_{\text{CUA}}^{\text{Tyr}}$ with the L-3-(2-naphthyl)alanine in the cell; and
- providing the cell with the L-3-(2-naphthyl)alanine, thereby producing in the cell at least one protein comprising the L-3-(2-naphthyl)alanine.

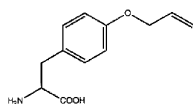
- 81.** A composition comprising an unnatural amino acid as shown in Formula I:

I



- 15 **82.** A composition comprising an unnatural amino acid as shown in Formula II:

II

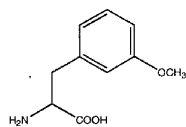


- 83.** A composition comprising an unnatural amino acid as shown in Formula III:

WO 02/085923

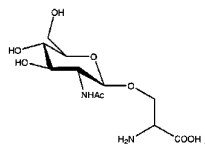
PCT/US02/12465

III



84. A composition comprising an unnatural amino acid as shown in Formula IV:

IV



- 5 85. The composition of claim 81, claim 82, claim 83, or claim 84, further comprising an orthogonal tRNA.
86. The composition of claim 85, wherein the unnatural amino acid is covalently bonded to the orthogonal tRNA.
87. The composition of claim 85, wherein the unnatural amino acid is covalently bonded to the orthogonal tRNA through an amino-acyl bond.
- 10 88. The composition of claim 85, wherein the unnatural amino acid is covalently bonded to a 3' OH or a 2' OH of a terminal ribose sugar of the orthogonal tRNA.
89. The composition of claim 81, claim 82, claim 83, or claim 84, further comprising an orthogonal aminoacyl tRNA synthetase.
- 15 90. The composition of claim 89, wherein the unnatural amino acid is hydrogen bonded to the orthogonal aminoacyl tRNA synthetase.
91. A protein comprising the unnatural amino acid of claim 81, claim 82, claim 83, or claim 84.

WO 02/085923

PCT/US02/12465

92. A cell comprising the unnatural amino acid of claim 81, claim 82, claim 83, or claim 84.

93. A method of synthesizing 4-allyl-L-tyrosine, the method comprising:

(a) reacting a protected tyrosine with allyl bromide, resulting in a protected allyl tyrosine; and,

(b) deprotecting the protected allyl tyrosine to produce 4-allyl-L-tyrosine.

94. The method of claim 93, comprising reacting the allyl bromide and the protected tyrosine in the presence of sodium hydride and DMF.

95. The method of claim 93, wherein the protected tyrosine comprises an NBoc protected tyrosine.

96. The method of claim 95, wherein (b) comprises deprotecting the protected allyl tyrosine in an acidic solution.

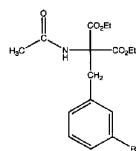
97. The method of claim 95, wherein (b) comprises deprotecting the protected allyl tyrosine in the presence of hydrochloric acid and dioxane.

98. The method of claim 93, further comprising extracting the 4-allyl-L-tyrosine with ethanol or dichloromethane.

99. A method of synthesizing a meta-substituted phenylalanine analog, the method comprising:

(a) condensing diethylacetamidomalonate and a meta-substituted benzyl bromide, resulting in a compound having formula V:

V



(b) hydrolyzing the compound having Formula V, resulting in a meta-substituted phenylalanine analog.

WO 02/085923

PCT/US02/12465

100. The method of claim **99**, wherein R comprises a keto, acetyl, or methoxy group.

101. The method of claim **99**, wherein the meta substituted phenylalanine analog comprises 3-methoxy-phenylalanine or 3-acetyl-phenylalanine.

5 **102.** The method of claim **99**, wherein the compound of Formula V comprises 1-(3-bromomethyl-phenyl)-ethanone.

103. The method of claim **99**, the method further comprising providing the meta substituted benzyl bromide by:

reacting N-bromosuccinimide with 3-methylacetophenone to produce a brominated
10 product; and,
crystallizing the brominated product in a hexane solution.

104. A method of preparing a glycosylated amino acid in vivo, the method comprising:

transforming a cell with a plasmid comprising a gene for an N-acetyl-
15 galactosaminidase, a transglycosylase, or a serine-glycosylhydrolase.

105. A method of preparing *p*-aminophenylalanine, the method comprising:

(a) enzymatically converting chorismate to 4-amino-4-deoxychorismic acid;

(b) enzymatically converting the 4-amino-4-deoxychorismic acid to 4-amino-4-deoxyprephenic acid;

20 (c) enzymatically converting the 4-amino-4-deoxyprephenic acid to *p*-aminophenyl-pyruvic acid:

(d) enzymatically converting the *p*-aminophenyl-pyruvic acid into *p*-aminophenylalanine.

106. The method of claim **105**, wherein (a) comprises contacting the chorismate
25 with a 4-amino-4-deoxychorismate synthase.

107. The method of claim **106**, wherein the 4-amino-4-deoxychorismate synthase comprises *papA*.

108. The method of claim **105**, wherein (b) comprises contacting the 4-amino-4-deoxychorismic acid with a chorismate mutase.

WO 02/085923

PCT/US02/12465

109. The method of claim 108, wherein the chorismate mutase comprises *papB*.
110. The method of claim 105, wherein (c) comprises contacting the 4-amino-4-deoxyprephenic acid with a prephenate dehydrogenase.
111. The method of claim 110, wherein the prephenate dehydrogenase
5 comprises *papC*.
112. The method of claim 105, wherein (d) comprises contacting the *p*-aminophenyl-pyruvic acid with an aminotransferase.
113. The method of claim 112, wherein the aminotransferase comprises a non-specific tyrosine aminotransferase derived from *E. coli*.
- 10 114. The method of claim 116, wherein the aminotransferase comprises *tyrB*, *aspS*, or *ilvE*.
115. The method of claim 105, wherein steps (a) through (d) are performed *in vivo*.
116. A method of producing *p*-aminophenylalanine in an *Escherichia coli* cell,
15 the method comprising transforming the cell with a plasmid comprising *papA*, *papB*, and *papC*, wherein the cell comprises chorismate and an aminotransferase and wherein expression of *papA*, *papB*, and *papC* results in a synthase, a mutase, and a dehydrogenase, wherein the synthase, the mutase, the dehydrogenase and the aminotransferase produce *p*-phenylalanine from chorismate.
- 20 117. A cell comprising:
- (a) a biosynthetic pathway system for producing an unnatural amino acid from one or more carbon sources within the cell; and,
- (b) a translation system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl tRNA synthetase (O-RS), wherein the O-RS preferentially
25 aminoacylates the O-tRNA with the unnatural amino acid and the O-tRNA incorporates the unnatural amino acid into a protein in response to a selector codon.
118. The cell of claim 117, wherein the cell comprises a bacterial cell.
119. The cell of claim 117, wherein the selector codon comprises a nonsense codon, a four base codon, or an amber codon.

WO 02/085923

PCT/US02/12465

120. The cell of claim 119, wherein the nonsense codon comprises a TAG codon.
121. The cell of claim 117, wherein the biosynthetic pathway system produces a natural cellular amount of the unnatural amino acid.
- 5 122. The cell of claim 117, wherein the biosynthetic pathway system comprises a synthase, a mutase, a dehydrogenase, and an aminotransferase.
123. The cell of claim 122, wherein the synthase comprises 4-amino-4-deoxychorismate synthase.
124. The cell of claim 122, wherein the mutase comprises chorismate mutase.
- 10 125. The cell of claim 122, wherein the dehydrogenase comprises prephenate dehydrogenase.
126. The cell of claim 122, wherein the synthase, the dehydrogenase, and the mutase are derived from *Streptomyces Venezuelae* or *Streptomyces pristinaespiralis* and the aminotransferase is derived from *E.coli*.
- 15 127. The cell of claim 117, wherein the unnatural amino acid comprises *p*-aminophenylalanine, dopa, or O-methyl-L-tyrosine.
128. The cell of claim 117, wherein the unnatural amino acid comprises a glycosylated amino acid or a pegylated amino acid.
129. The cell of claim 117, wherein the carbon source comprises chorismate.
- 20 130. The cell of claim 117, wherein the biosynthetic pathway system produces *p*-aminophenylalanine from chorsimate.
131. The cell of claim 117, wherein the biosynthetic pathway system:
- (a) enzymatically converts chorismate to 4-amino-4-deoxychorismic acid;
- (b) enzymatically converts 4-amino-4-deoxychorismic acid to 4-amino-4-deoxyprephenic acid;
- 25 (c) enzymatically converts 4-amino-4-deoxyprephenic acid to *p*-aminophenylpyruvic acid; and,
- (d) enzymatically converts the *p*-aminophenyl-pyruvic acid into *p*-aminophenylalanine.

WO 02/085923

PCT/US02/12465

132. The cell of claim 117, wherein the cell produces the unnatural amino acid in an amount sufficient for protein biosynthesis, which amount does not substantially alter a concentration of any natural amino acid or exhaust a cellular resource.

5 133. The cell of claim 132, wherein the unnatural amino acid is *p*-aminophenylalanine, L-dopa, or O-methyl-L-tyrosine.

134. The cell of claim 117, wherein the biosynthetic pathway system comprises a plasmid, which plasmid comprises a polynucleotide encoding a chorismate synthase, a polynucleotide encoding a chorismate mutase, and a polynucleotide encoding a prephenate dehydrogenase.

10 135. The cell of claim 117, wherein the cell comprises a plasmid, which plasmid comprises a *papA* gene, a *papB* gene, and a *papC* gene.

136. The cell of claim 135, wherein the plasmid further comprises an *lpp* promoter and a *lac* promoter.

15 137. The cell of claim 135, wherein the plasmid further comprises one or more ribosome binding site.

138. The cell of claim 135, wherein the plasmid is a low copy pSC101 derived plasmid.

20 139. A cell comprising one or more systems for producing at least twenty one amino acids and specifically incorporating one or more of the amino acids into one or more proteins within the cell, resulting in one or more incorporated amino acids, wherein at least one of the incorporated amino acids comprises an unnatural amino acid.

140. A method of identifying an advantage provided by an unnatural amino acid which has been incorporated into a cell, the method comprising:

25 (a) providing a library of cells, each of which cells comprises a randomized plasmid derived from an *E. coli* genome, wherein one or more of the randomized plasmids confers on the cells an ability to incorporate an unnatural amino acid into a protein.

30 (b) screening the library of cells to identify one or more cells with enhanced growth as compared to a native *E. coli* cell, thereby identifying an advantage provided by the unnatural amino acid.

WO 02/085923

PCT/US02/12465

1/37

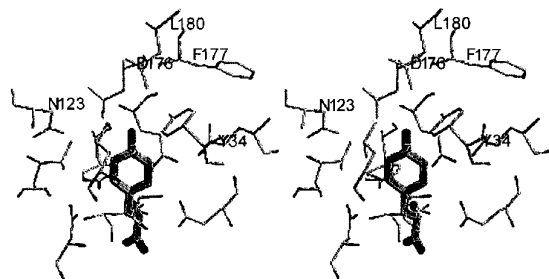


Fig. 1

WO 02/085923

PCT/US02/12465

2/37

Fig. 2A

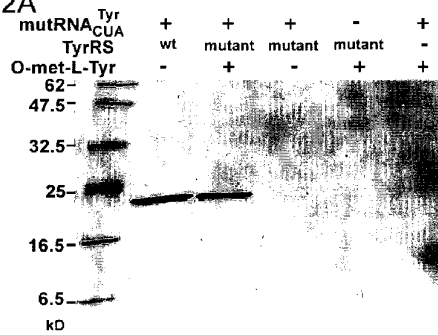


Fig. 2B



Fig. 2

WO 02/085923

PCT/US02/12465

3/37

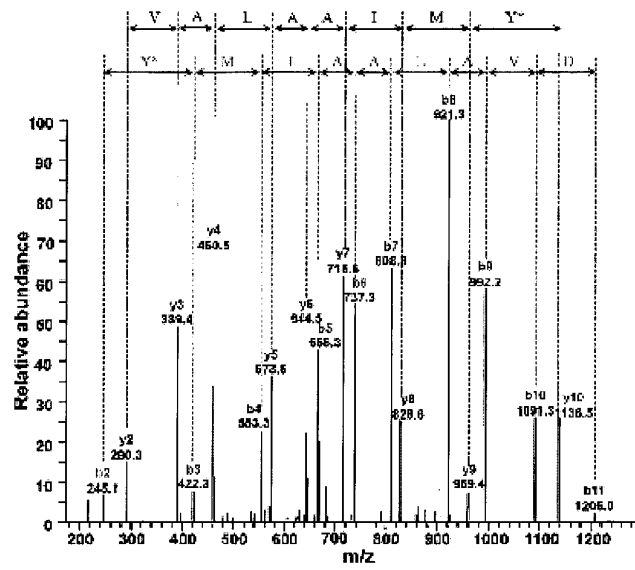


Fig. 3

WO 02/085923

PCT/US02/12465

4/37

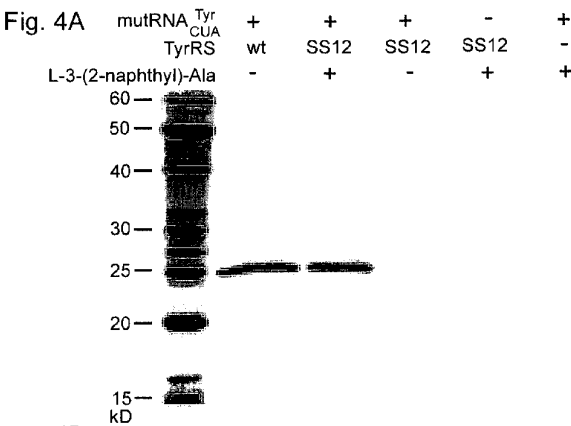


Fig. 4B



WO 02/085923

PCT/US02/12465

5/37

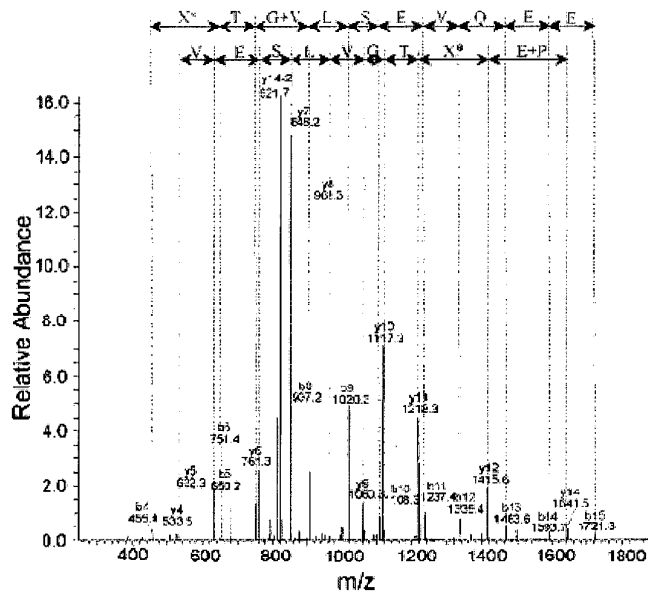


Fig. 5

WO 02/085923

PCT/US02/12465

6/37

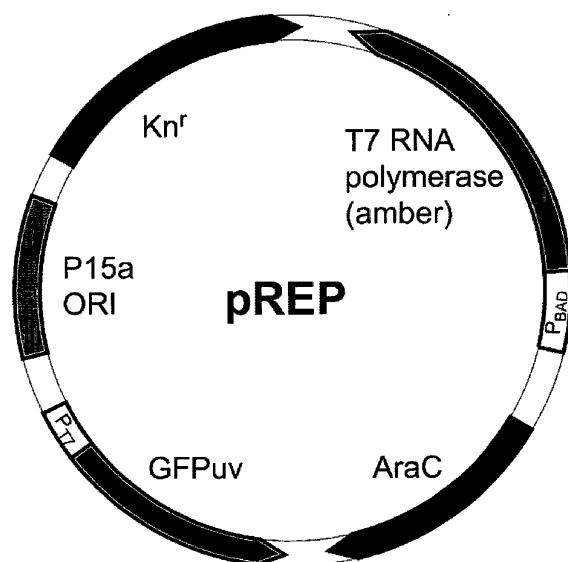


Figure 6A

WO 02/085923

PCT/US02/12465

7/37

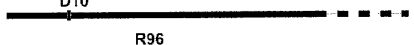
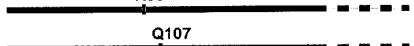
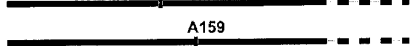
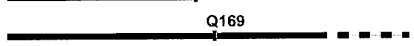
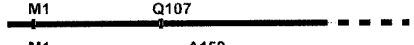
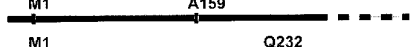
Construct Number		Fluorescence Enhancement (fold)
1		12
2		-
3		-
4		4
5		-
6		32
7		6
8		-
9		20
10		220
11		48
12		82

Figure 6B

WO 02/085923

PCT/US02/12465

8/37

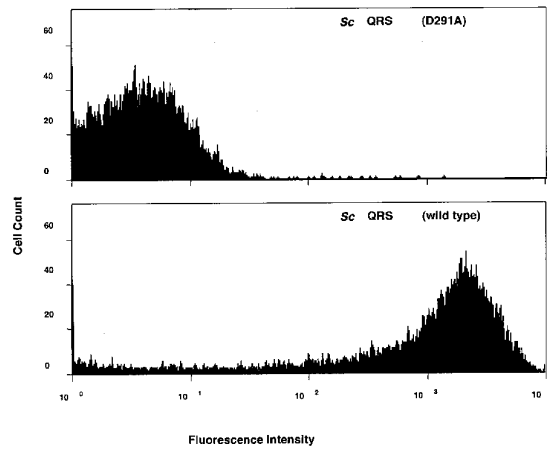


Figure 6C

WO 02/085923

PCT/US02/12465

9/37

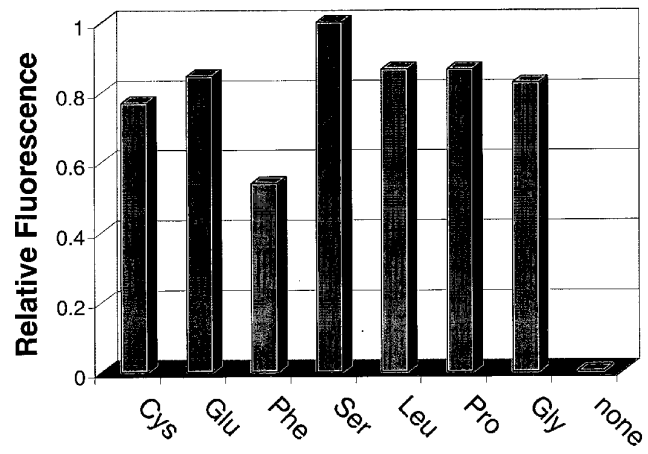


Figure 6D

WO 02/085923

PCT/US02/12465

10/37

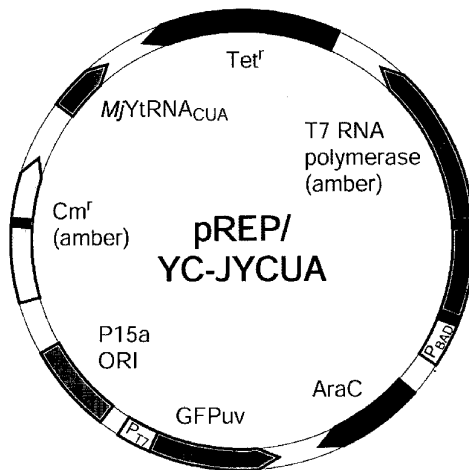


Figure 7A

WO 02/085923

PCT/US02/12465

11/37

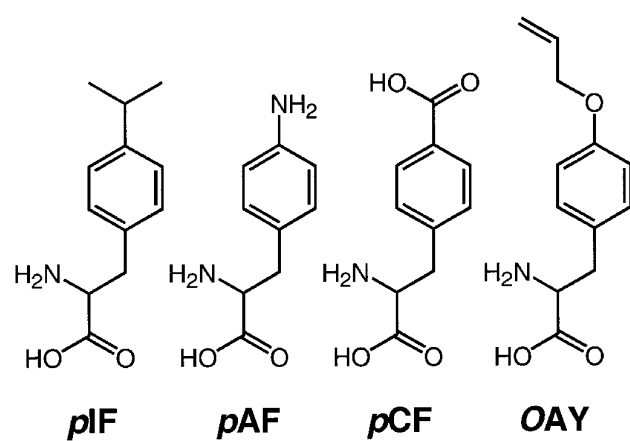


Figure 7B

WO 02/085923

PCT/US02/12465

12/37

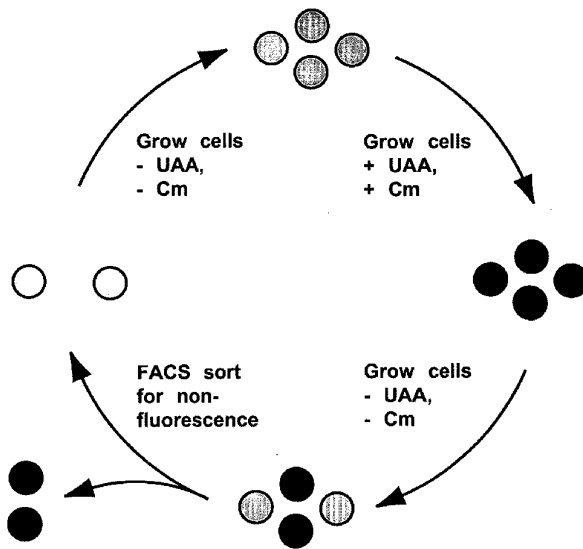
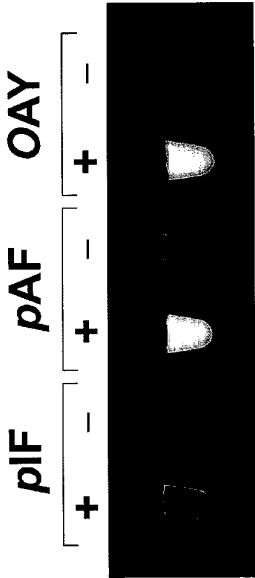


Figure 7C

13/37



WO 02/085923

PCT/US02/12465

14/37

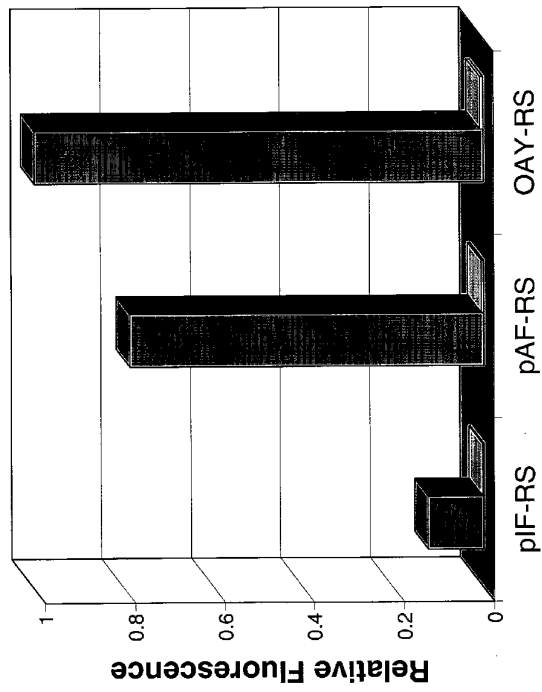


Figure 8B

15/37

Cm IC ₅₀ (μg/mL)	pIF-RS	pAF-RS	QAY-RS(1)
– Unnatural	< 5	< 5	< 5
+ Unnatural	75	100	120

Figure 8C

16/37

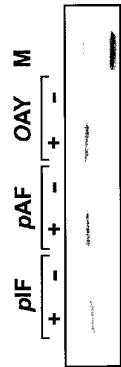


Figure 8D

WO 02/085923

PCT/US02/12465

17/37

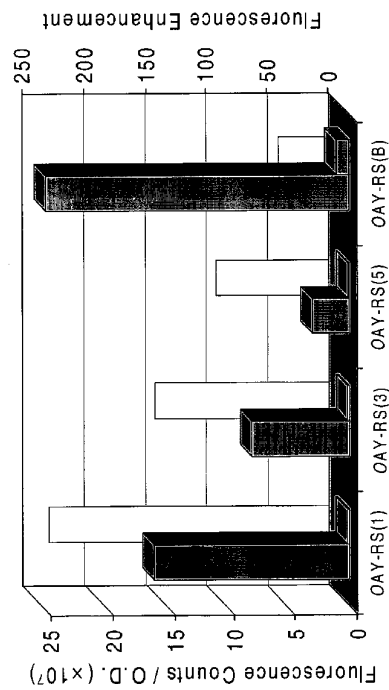


Figure 9

WO 02/085923

PCT/US02/12465

18/37

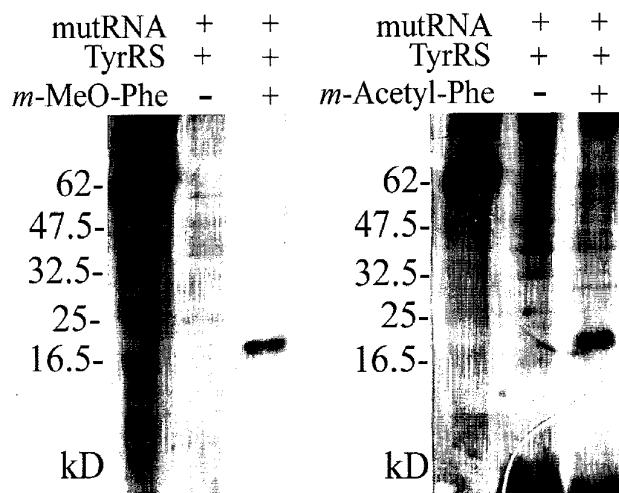


Figure 10

WO 02/085923

PCT/US02/12465

19/37

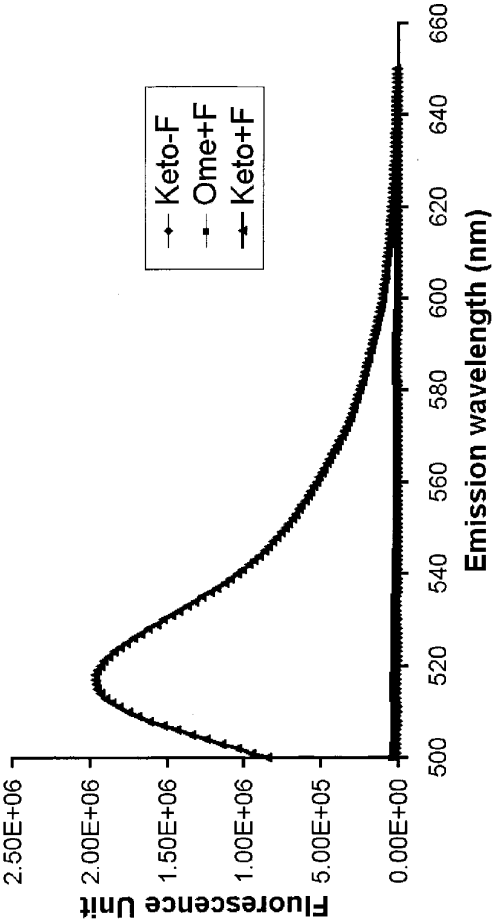


Figure 11

WO 02/085923

PCT/US02/12465

20/37

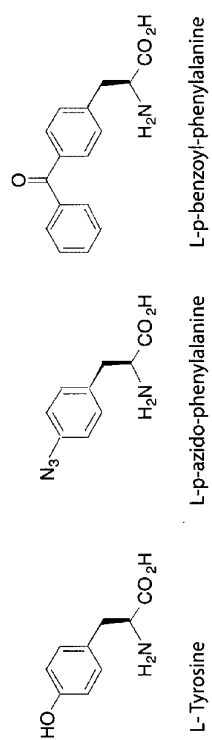


Figure 12

WO 02/085923

PCT/US02/12465

21/37

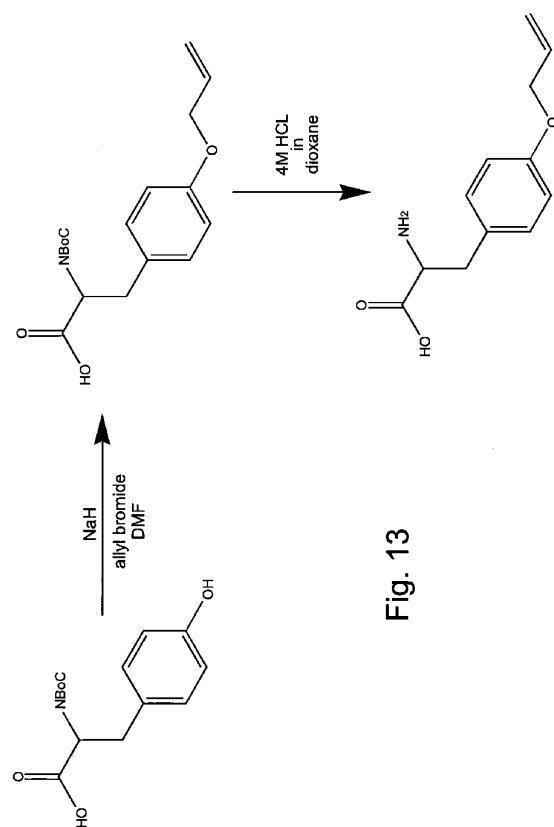


Fig. 13

WO 02/085923

PCT/US02/12465

22/37

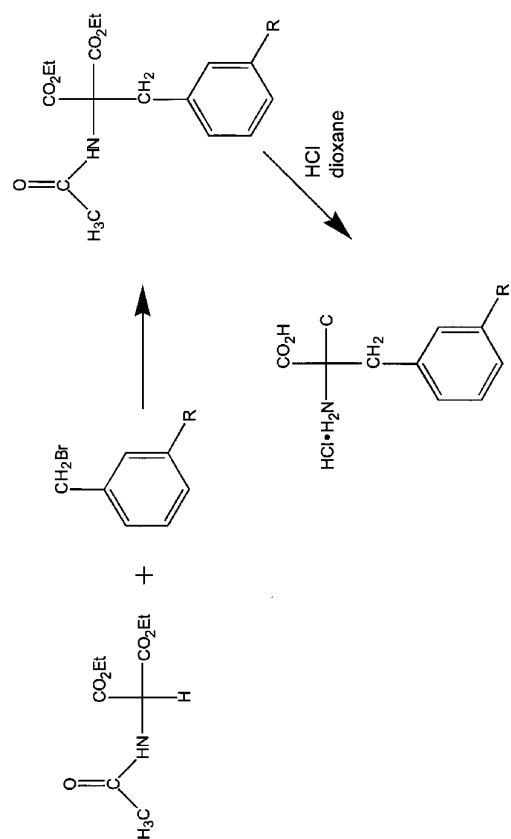


Fig. 14

Fig. 15A

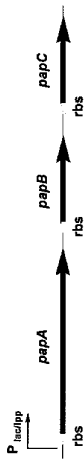
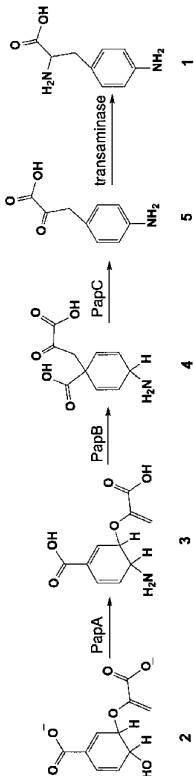


Fig. 15B



WO 02/085923

PCT/US02/12465

24/37

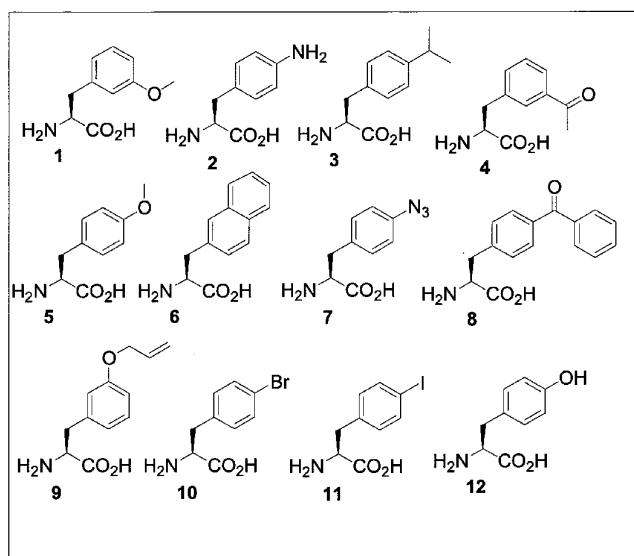


Fig. 16

WO 02/085923

PCT/US02/12465

25/37

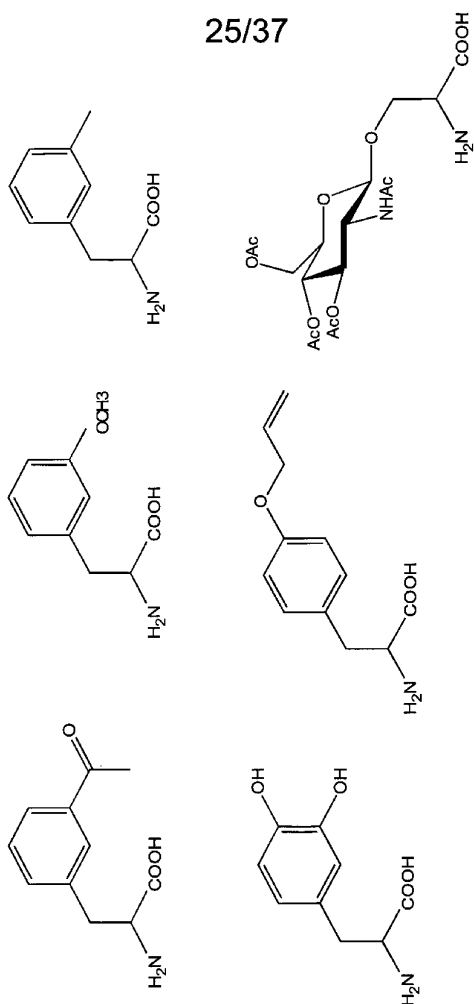


Fig. 17

26/37

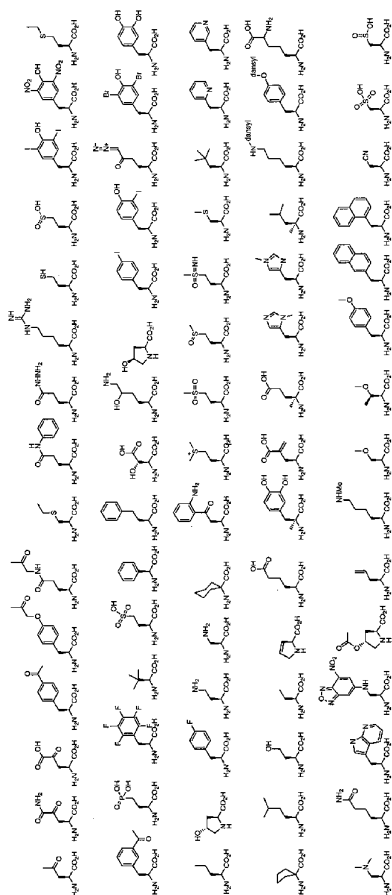


Fig. 18

27/37

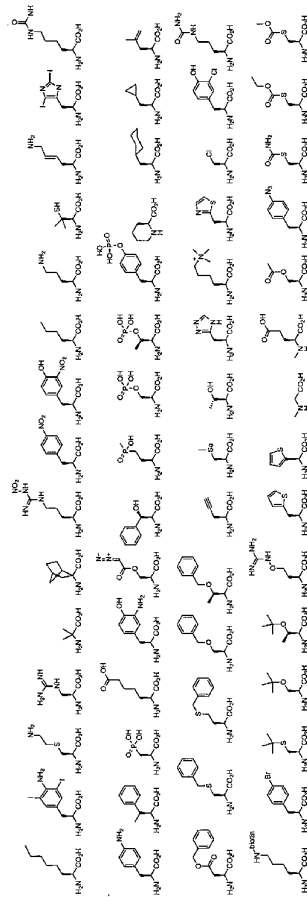


Fig. 19

WO 02/085923

PCT/US02/12465

28/37

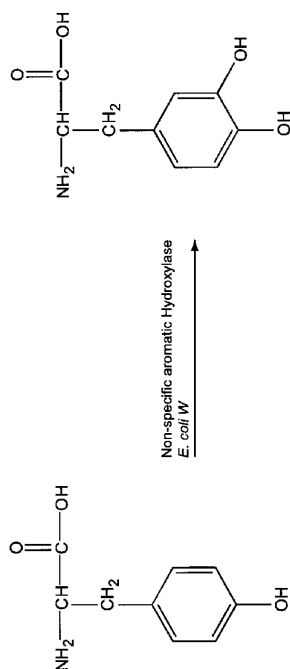


Fig. 20

29/37

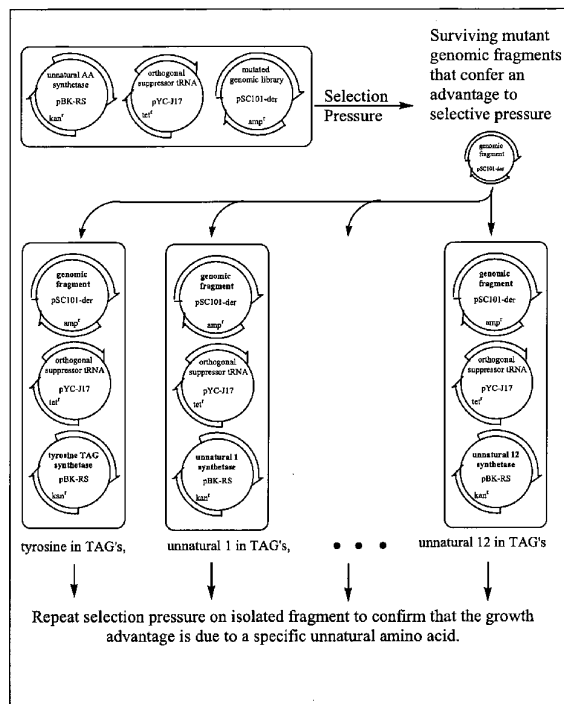


Fig. 21

30/37

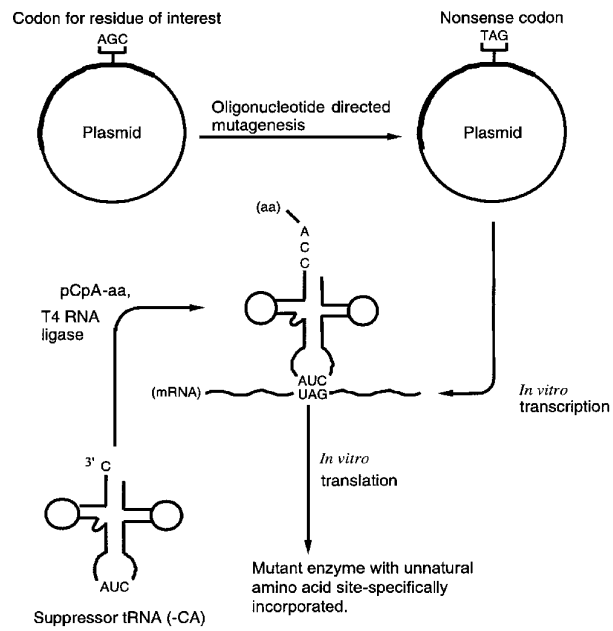


Fig. 22

WO 02/085923

PCT/US02/12465

31/37

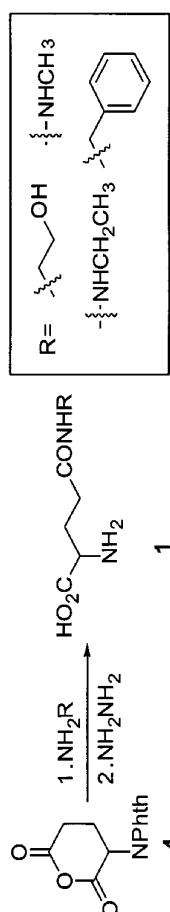


Fig. 23

WO 02/085923

PCT/US02/12465

32/37

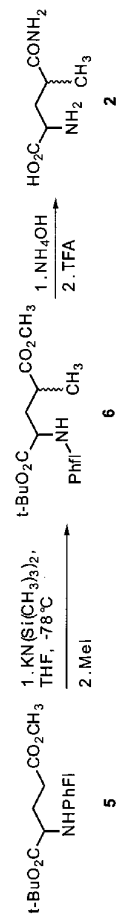


Fig. 24

WO 02/085923

PCT/US02/12465

33/37

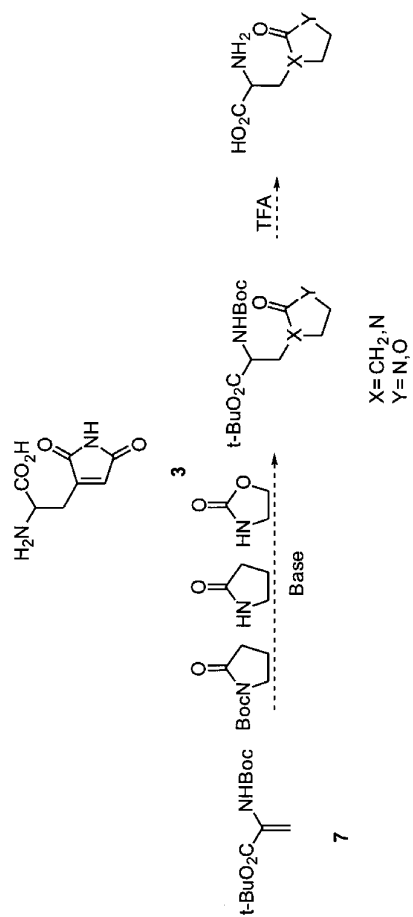


Fig. 25

WO 02/085923

PCT/US02/12465

34/37

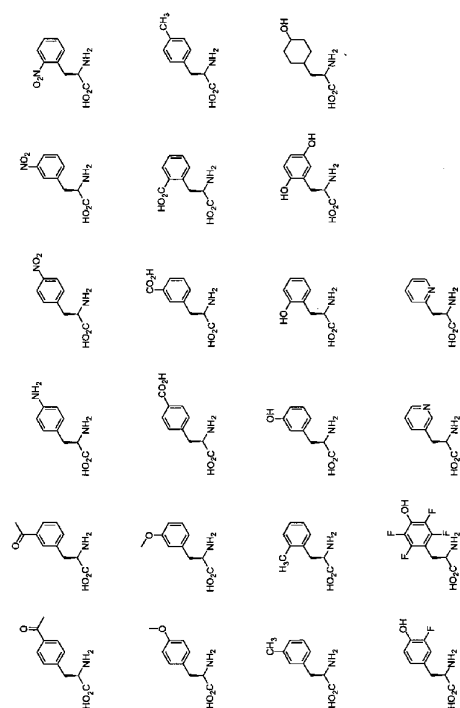


Fig. 26

WO 02/085923

PCT/US02/12465

35/37

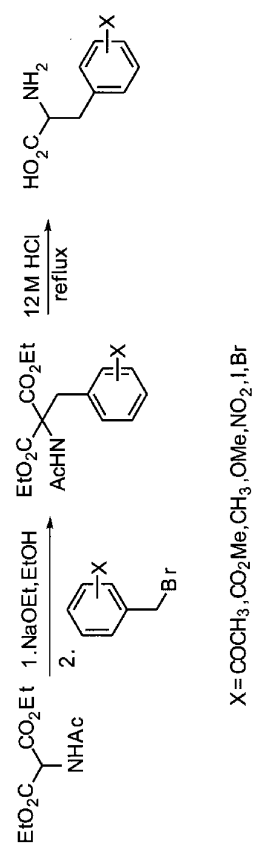


Fig. 27

WO 02/085923

PCT/US02/12465

36/37

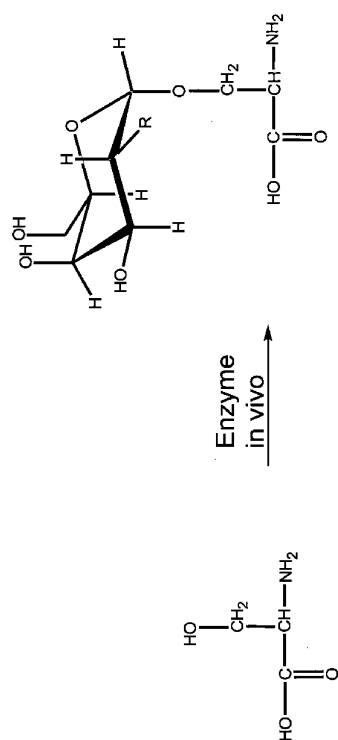


Fig. 28

WO 02/085923

PCT/US02/12465

37/37

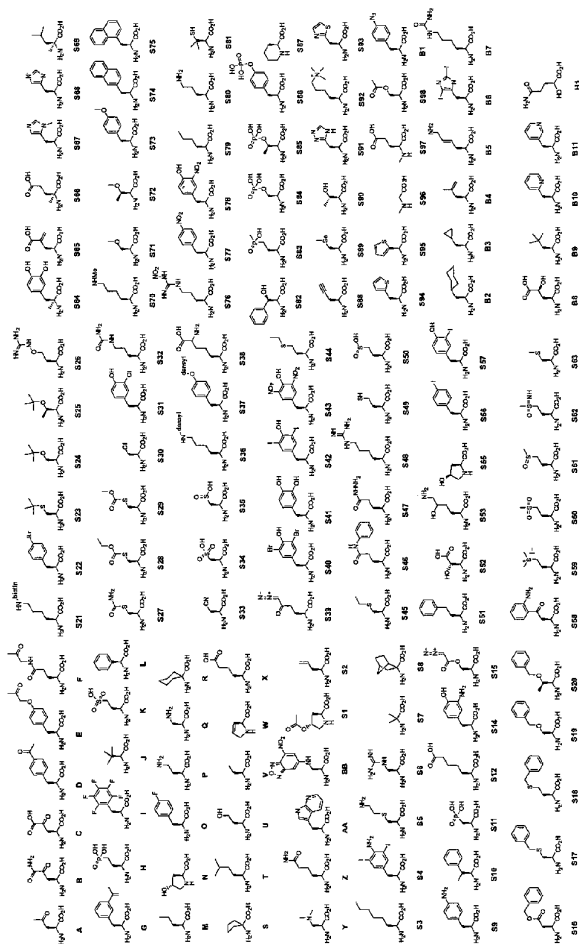


Fig. 29

【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
31 October 2002 (31.10.2002)

PCT

(10) International Publication Number
WO 2002/085923 A3

(51) International Patent Classification: C12N 9/00, 1/20,
1/14, 1/16, 1/18, C12P 21/06, C07H 21/04

[US/US]: 12668-63 Carmel Country Road, San Diego,
CA 92130 (US). **ZHANG, Zhiwen** [CN/US]: 12751 Via
Terceto, San Diego, CA 92130 (US).

(21) International Application Number:
PCT/US2002/012465

(74) Agents: **QUINE, Jonathan, Alan** et al.; Quine Intellectual
Property Law Group, P.C., P.O. Box 438, Alameda, CA
94501 (US).

(22) International Filing Date: 19 April 2002 (19.04.2002)

(25) Filing Language: English

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

(26) Publication Language: English

(30) Priority Data:
60/285,030 19 April 2001 (19.04.2001) US
60/355,514 6 February 2002 (06.02.2002) US

(71) Applicant (for all designated States except US): **THE
SCRIPPS RESEARCH INSTITUTE** [US/US], 10550
North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SCHULTZ, Peter**
[US/US], 1650 La Jolla Ranch Road, La Jolla, CA 92037
(US). **WANG, Lei** [CN/US], # 2320, 8730 Costa Verde
Blvd., San Diego, CA 92122 (US). **ANDERSON, John,
Christopher** [US/US], 3922 8th Avenue, Apt. 11, San
Diego, CA 92103 (US). **CHIN, Jason, W.** [GB/US], 4249
Nobel Drive #23, San Diego, CA 92122 (US). **LIU, David,
R.** [US/US], 1 Fox Run Lane, Lexington, MA 02420 (US).
MAGLIERY, Thomas, J. [US/US], 333 Mill Rd., North
Haven, CT 06473 (US). **MEGGERS, Eric, L.** [DE/US],
1534 Pine Street, Apt. 2F, Philadelphia, Pennsylvania
19102 (US). **MEHL, Ryan, Aaron** [US/US], 3576 31st
Street, San Diego, CA 92104 (US). **PASTRNAK, Miro**
[SK/US], 11686 Carmel Creek Road, Apt. 107, San
Diego, CA 92130 (US). **SANTORO, Stephen, William**

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments

(88) Date of publication of the international search report:
27 May 2004

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.



WO 2002/085923 A3

(54) Title: IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS

(57) Abstract: The invention provides methods and compositions for in vivo incorporation of unnatural amino acids. Also provided are compositions including proteins with unnatural amino acids.

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US02/12465															
A. CLASSIFICATION OF SUBJECT MATTER IPC(7) : C12N 9/00, 1/20, 1/14, 1/16, 1/18; C12P 21/06; C07H 21/04 US CL : 435/183, 252.3, 252.33, 254.11, 69.1; 536/24.1 According to International Patent Classification (IPC) or to both national classification and IPC																	
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 435/183, 252.3, 252.33, 254.11, 69.1; 536/24.1 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) C.Aplus																	
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category *</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>X,P --- Y,P</td> <td>WANG et al. A general approach for the generation of orthogonal tRNAs. Chemistry & Biology. September, 2001, Vol. 8, pages 883-890.</td> <td>1-4,6,10,12,13,15-18,48-51,53,66-68,70-75,78,79,117-125,127,129,132,133 ----- 7-9,32,34,64,65,77</td> </tr> <tr> <td>A</td> <td>LIU et al. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. Proc. Natl. Acad. Sci. 1997, Vol. 94, pages 10092-10097.</td> <td>1-9,16,17,48-54,64,65,72,73,77,78,117-121,132</td> </tr> <tr> <td>A</td> <td>LIU et al. Characterization of an 'orthogonal' suppressor tRNA derived from E. coli tRNA^{2Gln}. Chemistry & Biology. 1997, Vol. 4, pages 685-691.</td> <td>1-9,16,17,48-54,64,65,72,73,77,78,117-121,132</td> </tr> <tr> <td>A</td> <td>LIU et al. Progress toward the evolution of an organism with an expanded genetic code. Proc. Natl. Acad. Sci. 1999, Vol. 96, pages 4780-4785.</td> <td>1-9,16,17,48-54,64,65,72,73,77,78,117-121,132</td> </tr> </tbody> </table>			Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X,P --- Y,P	WANG et al. A general approach for the generation of orthogonal tRNAs. Chemistry & Biology. September, 2001, Vol. 8, pages 883-890.	1-4,6,10,12,13,15-18,48-51,53,66-68,70-75,78,79,117-125,127,129,132,133 ----- 7-9,32,34,64,65,77	A	LIU et al. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. Proc. Natl. Acad. Sci. 1997, Vol. 94, pages 10092-10097.	1-9,16,17,48-54,64,65,72,73,77,78,117-121,132	A	LIU et al. Characterization of an 'orthogonal' suppressor tRNA derived from E. coli tRNA ^{2Gln} . Chemistry & Biology. 1997, Vol. 4, pages 685-691.	1-9,16,17,48-54,64,65,72,73,77,78,117-121,132	A	LIU et al. Progress toward the evolution of an organism with an expanded genetic code. Proc. Natl. Acad. Sci. 1999, Vol. 96, pages 4780-4785.	1-9,16,17,48-54,64,65,72,73,77,78,117-121,132
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.															
X,P --- Y,P	WANG et al. A general approach for the generation of orthogonal tRNAs. Chemistry & Biology. September, 2001, Vol. 8, pages 883-890.	1-4,6,10,12,13,15-18,48-51,53,66-68,70-75,78,79,117-125,127,129,132,133 ----- 7-9,32,34,64,65,77															
A	LIU et al. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. Proc. Natl. Acad. Sci. 1997, Vol. 94, pages 10092-10097.	1-9,16,17,48-54,64,65,72,73,77,78,117-121,132															
A	LIU et al. Characterization of an 'orthogonal' suppressor tRNA derived from E. coli tRNA ^{2Gln} . Chemistry & Biology. 1997, Vol. 4, pages 685-691.	1-9,16,17,48-54,64,65,72,73,77,78,117-121,132															
A	LIU et al. Progress toward the evolution of an organism with an expanded genetic code. Proc. Natl. Acad. Sci. 1999, Vol. 96, pages 4780-4785.	1-9,16,17,48-54,64,65,72,73,77,78,117-121,132															
☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.																	
<table border="1"> <thead> <tr> <th>* Special categories of cited documents:</th> <th>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</th> </tr> </thead> <tbody> <tr> <td>"A" documents defining the general state of the art which is not considered to be of particular relevance</td> <td>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</td> </tr> <tr> <td>"E" earlier application or patent published on or after the international filing date</td> <td>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</td> </tr> <tr> <td>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td> <td>"&" document member of the same patent family</td> </tr> <tr> <td>"O" document referring to an oral disclosure, use, exhibition or other means</td> <td></td> </tr> <tr> <td>"P" document published prior to the international filing date but later than the priority date claimed</td> <td></td> </tr> </tbody> </table>			* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	"A" documents defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	"E" earlier application or patent published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family	"O" document referring to an oral disclosure, use, exhibition or other means		"P" document published prior to the international filing date but later than the priority date claimed				
* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention																
"A" documents defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone																
"E" earlier application or patent published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art																
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family																
"O" document referring to an oral disclosure, use, exhibition or other means																	
"P" document published prior to the international filing date but later than the priority date claimed																	
Date of the actual completion of the international search 27 January 2004 (27.01.2004)		Date of mailing of the international search report 25 MAR 2004															
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230		Authorized officer Kathleen M Kerr <i>Kathleen M Kerr</i> Telephone No. (571) 272-1600															

INTERNATIONAL SEARCH REPORT		PCT/US02/12465
C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KOWAL et al. Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and eubacteria. Proc. Natl. Acad. Sci. February, 2001, Vol. 98, No. 5, pages 2268-2273.	1-9,16,17,48-54,64,65,72,73,77,78,117-121,132

INTERNATIONAL SEARCH REPORT		International application No. PCT/US02/12465
Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)		
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:		
1.	☐	Claim Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	☒	Claim Nos.: 11 and 76 because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically: Please See Continuation Sheet
3.	☐	Claim Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)		
This International Searching Authority found multiple inventions in this international application, as follows: Please See Continuation Sheet		
1.	☐	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.	☐	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	☐	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.	☒	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-10, 12-19, 48-54, 64-75, 77-80, and 117-138
Remark on Protest	☐	The additional search fees were accompanied by the applicant's protest.
	☐	No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/12465

Continuation of Box I Reason 2:

Claims 11 and 76 are unsearchable because the claims are too confusing for a meaningful search to be carried out. Both claims refer to SEQ ID NOs: 4-34 as polypeptide sequences; however, these sequences are nucleic acid sequences as filed on February 21, 2003. Moreover, both claims refer to SEQ ID NOs: 35-66 as nucleic acid sequences; however, these sequences are polypeptide sequences as filed on February 21, 2003. Thus, without any clear subject matter to be examined, these claims are considered unsearchable in their entirety.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-10, 12-19, 48-54, 64-75, 77-80, and 117-138, drawn to a composition of a translation system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl tRNA synthetase (O-RS), examples thereof in *E. coli* cells, cells that contain the translation system plus a biosynthetic pathway for producing an unnatural amino acid, and methods of using said translation system.

Group II, claim(s) 20-47 and 55-63, drawn to proteins containing unnatural amino acids.

Group III, claim(s) 81 and 85-92, drawn to an unnatural amino acid of Formula I, proteins, cells, orthogonal aminoacyl tRNA synthetases and orthogonal tRNAs thereof.

Group IV, claim(s) 82 and 85-92, drawn to an unnatural amino acid of Formula II, proteins, cells, orthogonal aminoacyl tRNA synthetases and orthogonal tRNAs thereof.

Group V, claim(s) 83 and 85-92, drawn to an unnatural amino acid of Formula III, proteins, cells, orthogonal aminoacyl tRNA synthetases and orthogonal tRNAs thereof.

Group VI, claim(s) 84-92, drawn to an unnatural amino acid of Formula IV, proteins, cells, orthogonal aminoacyl tRNA synthetases and orthogonal tRNAs thereof.

Group VII, claim(s) 93-98, drawn to methods of making 4-allyl-L-tyrosine.

Group VIII, claim(s) 99-103, drawn to methods of making a meta-substituted phenylalanine analog.

Group IX, claim(s) 104, drawn to methods of preparing a glycosylated amino acid *in vivo*.

Group X, claim(s) 105-116, drawn to methods of making p-aminophenylalanine using enzymes.

Group XI, claim(s) 139, drawn to cells that make and incorporate at least twenty one amino acids.

Group XII, claim(s) 140, drawn to methods that identify an advantage to incorporating an unnatural amino acid.

The inventions listed as Groups I-XII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons.

The special technical feature of Group I is the combination of the structure and function of the products in the translation system that include both an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl tRNA synthetase (O-RS). The structure of these products is either defined by homology with the art or by particular species in Claim 10; the function of these products is both defined in the art and in Claim 1, particularly with the phrase "preferentially aminoacylates". Claims 18-19 are particular species of the genus of Claim 1, wherein a species of the translation system is found in a particular cell. Claims 48-54, 64-75, and 77-78 are a first method of using the special technical feature of the products of the first Group; Claims 79-80 are species of methods of using the special technical feature of the products of the first Group. Claims 117-138 require the translation system of Claim 1 and additional components, thus sharing the special technical feature of Group I. Thus, all the members of Group I are products requiring the special technical feature of Group I or are methods of using these products, wherein the special technical feature is the combination of the art-defined and species-defined structures and the function of preferentially aminoacylating with an unnatural amino acid.

INTERNATIONAL SEARCH REPORT

PCT/US02/12465

The proteins of Group II do not share the special technical feature of Group I, having neither an O-tRNA nor an O-RS or their structural and/or functional equivalents. Moreover, the proteins of Group II do not have a special technical feature within their Group as noted in the species election below.

The unnatural amino acids of Groups III-VI do not share the special technical feature of Group I, having neither an O-tRNA nor an O-RS or their structural and/or functional equivalents. Moreover, in Claims 89 and 90 wherein an O-RS is required in addition to the unnatural amino acid, the functional equivalents required as part of the special technical feature of Group I are not met having no requirement for preferentially aminoacylating with the attached unnatural amino acid.

The methods of Groups VII-X and XII do not share the special technical feature of Group I, using neither an O-tRNA nor an O-RS or their structural and/or functional equivalents.

The cells of Group XI do not share the special technical feature of Group I, requiring neither an O-tRNA nor an O-RS or their structural and/or functional equivalents. While the translation system of Group I can be used in the cells as a species of the claimed genus, the genus of Group XI is much more broad than requiring this special technical feature in either structural or functional terms.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack unity of invention because they are not so linked as to form a single general inventive concept under PCT Rule 13.1.

In order for more than one species to be examined, the appropriate additional examination fees must be paid. The species are as follows:
In Group II, the type of protein that contains an unnatural amino acid are individual species.

The claims are deemed to correspond to the species listed above in the following manner:

Claims 21, 22, 24, 30, 31, 32, 38, 44, 45, 47, 56, 57, and 59 list species of proteins containing unnatural amino acids.
The following claim(s) are generic: 20, 23, 25-29, 33-37, 39-43, 58, and 60-63.

The species listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: proteins with unnatural amino acids are not novel in the art (see Nakagawa *et al.* Biochemistry (2000) 39:15826-15835).

フロントページの続き

(51)Int.Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 1/15	C 1 2 N 1/21	4 H 0 4 5
C 1 2 N 1/19	C 1 2 P 13/04	
C 1 2 N 1/21	C 1 2 P 21/02	C
C 1 2 N 5/10	C 1 2 Q 1/06	
C 1 2 P 13/04	C 1 2 N 5/00	A
C 1 2 P 21/02	A 6 1 K 37/02	
C 1 2 Q 1/06	A 6 1 K 37/24	

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PL,PT,R O,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

- (72)発明者 レイ・ワン
アメリカ合衆国 カリフォルニア州 9 2 1 2 2 サン ディエゴ コスタ ヴァード ブールヴァード 8 7 3 0 ナンバー 2 3 2 0
- (72)発明者 ジョン・クリストファー・アンダーソン
アメリカ合衆国 カリフォルニア州 9 2 1 0 3 サン ディエゴ アpartment 1 1 エイス アヴェニュー 3 9 2 2
- (72)発明者 ジェイソン・ダブリュー・・チン
アメリカ合衆国 カリフォルニア州 9 2 1 2 2 サン ディエゴ ノーベル ドライヴ ナンバー 2 3 4 2 4 9
- (72)発明者 デイヴッド・アール・・リウ
アメリカ合衆国 マサチューセッツ州 0 2 4 2 0 レキシントン フォックス ラン レーン 1
- (72)発明者 トーマス・ジェー・・マグライリー
アメリカ合衆国 コネティカット州 0 6 4 7 3 ノース ヘヴン ミル ロード 3 3 3
- (72)発明者 エリック・エル・・メッゲルス
アメリカ合衆国 ペンシルベニア州 1 9 1 0 2 フィラデルフィア アpartment 2 エフ パイン ストリート 1 5 3 4
- (72)発明者 ライアン・アaron・メール
アメリカ合衆国 カリフォルニア州 9 2 1 0 4 サン ディエゴ サーティファースト ストリート 3 5 7 6
- (72)発明者 ミロ・パストラナック
アメリカ合衆国 カリフォルニア州 9 2 1 3 0 サン ディエゴ アpartment 1 0 7 カーメル クリーク ロード 1 1 6 8 6
- (72)発明者 ステファン・ウィリアム・サントロ
アメリカ合衆国 カリフォルニア州 9 2 1 3 0 サン ディエゴ カーメル カントリー ロード 1 2 6 6 8-6 3
- (72)発明者 ツィーウェン・チャン
アメリカ合衆国 カリフォルニア州 9 2 1 3 0 サン ディエゴ ヴィア テルセト 1 2 7 5 1

Fターム(参考) 4B024 AA01 AA03 AA11 AA20 CA01 CA11 DA06 GA11 HA20
4B063 QA05 QQ05 QS24 QX01
4B064 AE03 AG18 CA02 CA19 CC24 DA01 DA13 DA20
4B065 AA26X AB01 AC14 BA02 BB12 BB14 CA17 CA24 CA44 CA46

4C084	AA02	AA07	DB56	MA01	NA05	ZA512	
4H045	AA10	AA20	BA50	CA40	DA13	EA20	FA74